Three Essays in Financial Economics

Dissertation
for the Faculty of Economics, Business Administration
and Information Technology of the University of Zurich

to achieve the title of
Doctor of Philosophy
in Banking & Finance

presented by

Urs Schweri
from Siglistorf AG

approved in October 2010 at the request of
Prof. Dr. Thorsten Hens
Prof. Dr. Alexandre Ziegler

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorises the printing of this Doctoral Thesis, without thereby giving any opinion on the views contained therein.

Zurich, 27 October 2010
Chairman of the Doctoral Committee: Prof. Dr. Dieter Pfaff

Acknowledgments

I would like to thank my supervisor Thorsten Hens for helpful suggestions, inspiring ideas and his continuous guidance and Alexandre Ziegler and Enrico De Giorgi for their unstinting support. I am especially grateful to my coauthor Klaus Reiner Schenk-Hoppé for his perceptive questions and brilliant insights and to Hermann Garbers, my former econometrics professor and the sponsor of the first year of my PhD studies. For their interesting discussions and useful observations, I thank Wolfgang Härdle, Stefan Reimann, Marc Oliver Rieger, Alexander Wagner, Peter Wöhrmann, Gorazd Brumen, Nilufer Caliskan, Rouslan Moro, Christian Reichlin and Andreas Tupak. Finally, I would like to thank Martine Baumgartner, the secretary of the chair, for her administrative and moral support.

I gratefully acknowledge the financial assistance received from the National Center of Competence in Research 'Financial Valuation and Risk Management' (NCCR FinRisk) managed by the Swiss National Science Foundation on behalf of the federal authorities. My research visit to Humbold University in Berlin was funded by the Collaborative Research Center 649: Economic Risk. The data for this work were provided by the following institutions: Research Priority Program Finance and Financial Markets of the University of Zurich, the Department of Banking and Finance of the University of Zurich and the Swiss Finance Institute.

Lastly and most importantly, I am indebted to my family, friends and colleagues who supported me during my research.

Introduction

The concept of absence of arbitrage is a cornerstone of modern finance. It implies that financial gain cannot be derived from nothing or in simpler terms, an asset with higher expected returns is riskier. The no-arbitrage condition implies the existence of a pricing kernel that precisely reflects the pricing information of all assets. From the stipulation, it follows that the price of an asset is equal to its pricing kernel-weighted payoffs. In fact, the realizations of the pricing kernel vary in response to economic situations such as a recession or a boom. Typically, during a recession, when the value of money is higher, the pricing kernel has a higher value than during a boom. Classical finance theories such as the capital asset pricing model (CAPM) equate good and bad economic conditions with high and low equity market returns respectively. This, in turn, implies that the pricing kernel is declining in market returns. However, estimations of higher-order moments in CAPM models on equity returns show that the pricing kernel is U-shaped; see Dittmar (2002), Potì (2006) and Post et al. (2008), that is, it increases with positive returns. This finding thus contradicts the classical finance theory.

The first article mainly investigates whether the upward slope in the U-shaped pricing kernel is merely a statistical artifact. To determine this, different functional forms of the pricing kernel are estimated on the basis of the generalized method of moments (GMM). The estimated polynomial pricing kernel for three of the five datasets is U-shaped with a clear increasing region. In these three datasets, the observed average asset returns are better explained by the U-shaped pricing kernels. This shape is robust for different time windows. Instead of polynomials, estimating the kernels with other functional forms that do not exclude a U-shape by construction also yields a U-shaped kernel. However, because the upward slopes are not statistically significant, these findings may not succeed in persuading advocates of standard economic theory to change their viewpoints.

In the first article, asset returns are assumed to be exogenous. In reality, though, assets are never wholly intangible; they are backed by real companies that grow and evolve with time. Empirical evidence shows that firm characteristics change throughout the various phases of evolution that is, they follow a life cycle. Young firms are more prone to risk. They encounter greater challenges in gaining access to credit markets and therefore often reinvest their earnings. At the same time, they place greater emphasis on innovation and have a faster growth curve. Large companies, however, tend to be risk-averse, less flexible, and slower in growth. However, because of their slower pace of growth, they have free cash flows to pay out as dividends. The second article captures these observations and presents them in a simple life cycle model. The following assumptions generate the described life cycle features in the model: decreasing marginal productivity, that is, an inverse relationship between firm capital and productivity, relative to the firm size; over proportional negative exogenous random shocks to small firms; and lack of access to external finance sources. Overproportional shocks to small firms result in a complex relationship between firm size, expected return and volatility. With larger firms, a so-called leverage effect can be observed: a drop in the value of a company increases the volatility of the returns of it. With small and medium-sized firms, a value effect is observed: on an average, companies with high book values yield higher returns because the high book value helps absorb the negative shocks. Naturally, the higher chance of survival also makes companies with more capital more valuable. With the help of business cycles, the life cycle model also explains IPO waves, procyclical investment behavior, and countercyclical default probabilities.

The second article also explores whether intervention by a central bank can help the small constrained firms. Findings reveal that in the absence of access to external finance, an interest rate policy introduced by the central bank offers little assistance. The interest rate policy has negligible impact on the buffer against negative shocks that constrained firms attempt to build by saving a considerable part of the capital. Consequently, the investment behavior remains unchanged. The only remedial measure for these constrained companies is acquiring access to the credit markets. However, agency conflicts are one of the main reasons that prevent many companies from securing sufficient credit, and in most cases, they cannot be easily resolved through regulations.

The last article employs the dividend process from the second article in an evolutionary market model. In this model, several investment strategies start with a sum of initial wealth and compete against each other. The study aims to identify the strategy that can not only survive but also outperform the other strategies in the market. The model assumes three dividend-based stages of life cycles. IPOs are essentially newly founded firms that pay no dividends. They evolve into startups, which are small firms, in the subsequent period and pay a few dividends. Startups face a higher risk of default but also stand to grow into a concern. Concerns, on the other hand, pay out large dividends and face a low default risk. Every economic period witnesses the birth of new IPOs and defaults. Different investment strategies compete in an evolutionary market with these three types of assets. High-performing strategies naturally gain more wealth shares than low-performing ones. The success of a strategy depends not only on its ability to predict the dividends, which follow a life cycle, but also on the other strategies in the market.

A surviving strategy for infinitely lived assets is a generalized Kelly rule, a strategy that maximizes the long-term growth of wealth. Simulation results from the third article, which also discusses non-infinitely lived assets following life cycles, are in agreement with this finding. A novel finding of this article is that strategies that predict the dividend process of a firm on the basis of those of other similar firms perform better and gain more wealth shares. Because events such as defaults occur once in the lifetime of a firm, past realization of the dividends of that firm reveal almost nothing about a default. Lack of sufficient data history for predicting the dividend process can lead to a situation wherein simplistic strategies perform better than an estimated generalized Kelly rule. Given that a default is a rare event, it takes considerably long for the generalized Kelly rule to dominate the market.

Bibliography

Robert F. Dittmar. Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns. *Journal of Finance*, 57(1):369–403, February 2002.

Thierry Post, Pim van Vliet, and Haim Levy. Risk Aversion and Skewness Preference. *Journal of Banking & Finance*, 32(7):1178–1187, July 2008.

Valerio Potì. The Coskewness Puzzle in the Cross-Section of Industry Portfolio Returns. Working papter, Dublin City University Business School, March 2006.

Contents

Article 1

Is the Pricing Kernel U-Shaped?

Abstract: There is strong empirical evidence that the pricing kernel is U-shaped, which provides a way to explain the substantial coskewness premium. Existing studies typically use a polynomial approximation of the pricing kernel. Problematically, these polynomials have, in most cases, increasing parts by construction. Therefore, it is not clear whether the increasing parts are an artifact of the chosen functional form. Taking this concept into consideration, this paper shows that pricing kernels, as estimated by the generalized method of moments on equity data, are still U-shaped and that the increasing part is not a statistical artifact. This conclusion derives from the fact that the functional form of kernels, which allows for strictly decreasing kernels as well as for kernels with increasing parts, is still U-shaped. These results arise from checking for higher order polynomials, various time horizons, and different functional forms of the kernel.

1.1 Introduction

In the absence of arbitrage, a pricing kernel exists such that the price of an asset is equal to its pricing kernel-weighted payoffs. However, the no-arbitrage condition provides no information about the shape of the kernel (besides that of non-negativity). In equilibrium models with complete markets and a risk-averse representative agent who knows the probabilities of all states of the world, the pricing kernel is high in states with low resources because the marginal utility of one unit of additional consumption is high. However, in states with many resources, the pricing kernel is low. Therefore, the pricing kernel should decrease with resources.[1] In contradiction, there is, as shown in the following paragraphs, empirical evidence that the kernel is U-shaped within a certain range. However, studies such as those from Dittmar (2002), Potì (2006), and Post et al. (2008) do not investigate whether these increasing parts are significant. This paper tries to fill this gap.

Empirical evidence for increasing parts in the pricing kernel can be found in equity as well in option data. Estimating the kernel with equity returns yields a U-shaped pricing kernel, as Dittmar (2002), Potì (2006), and Post et al. (2008) have shown by approximating the kernel with a quadratic function or a higher order polynomial. The reason underlying this shape is the coskewness of single assets with the market portfolio returns (as a proxy for the available resources), such that the three-moment extension of the capital asset pricing model (CAPM) provides a significant risk premium for coskewness (Kraus and Litzenberger (1976), Friend and Westerfield (1980), Barone-Adesi (1985), Lim (1989), Harvey and Siddique (2000), Errunza and Sy (2005) and Smith (2007)). However, Dittmar (2002) and Post et al. (2008) show that the observed coskewness premium can no longer be explained if nonsatiation, risk aversion, and nonincreasing absolute risk aversion are imposed on the utility function of the representative agent (thereby translating into restrictions on the pricing kernel). Furthermore, Potì and Wang (2010) showed that unconstrained quadratic or higher order kernels imply relative risk aversions of above five for the representative investor, which is generally considered to be implausible. The pricing kernel in the well-known CAPM, discussed by Sharpe (1964), Lintner (1965), and Mossin (1966), is linear in relation to market returns, whereas the three-moment CAPM extends this with a quadratic term. On extending the CAPM by further moments, Fang and Lai (1997) and Hung (2008) found that co-kurtosis is also a relevant pricing factor. Considering a third order polynomial for the pricing, it still remains U-shaped in market returns,

[1] Suitable textbook references are Magill and Quinzii (1996) or Cochrane (2001).

as shown by Dittmar (2002). Estimations of the kernel from equity data clearly point to a U-shaped kernel.

Pricing kernel estimations done with option data also show a U-shaped kernel around the current stock price. Prominent examples of this are Aït-Sahalia and Lo (2000), Jackwerth (2000), and Rosenberg and Engle (2002). Note that estimating pricing kernels with option data allows us to estimate the pricing kernel within a broader range of market portfolio returns than is the case for equity data. The reason for this is that options exist with quite extreme strike prices, whereas such extreme values can rarely be observed in markets. Therefore, it turns out that pricing kernels typically fall after an increasing part of the kernel. If the view is restricted to the range of the kernel, which can be estimated by equity data, the kernel estimated by option data has a U-shape. More recent evidence is less clear: Detlefsen et al. (2007) estimated the kernel for German data at several points in time; in some cases, he derived a U-shape around the actual index level, and sometimes the kernel was merely decreasing. In another work, Golubev et al. (2008) provide some evidence that the increasing parts of kernels estimated using option prices are statistically significant. However, by using an asymmetric GJR-GARCH model with empirical innovations for option data, Barone-Adesi et al. (2008) and Barone-Adesi and Dall'O (2010) showed that the increasing parts of the kernel largely disappear.

In conclusion, there is empirical evidence for U-shaped kernels from equity as well as from option data around the current market price. Given that the estimations with equity data typically employ polynomial kernels and particularly quadratic kernels, the pricing kernel, by construction, has to be U-shaped (an inverse U-shape is also conceivable). This leads to the question as to whether the increasing part of the pricing kernel is merely a consequence of the polynomial functional form. Another interesting question is whether many observations of market returns are present in the increasing part. If not, then the increasing part could be a meaningless artifact. Unfortunately, the existing literature is not particularly helpful in addressing these concerns. For instance, Dittmar (2002), Potì (2006) and Post et al. (2008) estimated kernels with increasing parts and found that the coefficients before the polynomial terms were significant. However, their studies did not locate the minimum of the kernel or the origin of the increasing part of the kernel. Further, they do not provide any information on how often the economy lies in the increasing part of the kernel. Other studies, such as Hansen and Singleton (1982, 1983), and many subsequent works, have chosen a functional form (typically a power utility) such that the pricing kernel declines with any parameterization. More recently, Post and van Vliet (2006) and De Giorgi and Post (2008) estimated pricing kernels based on second-order stochastic dominance—and by that, assuming decreasing kernels—and found that the pricing kernels are steep in losses and flat in gains. However, this approach is also unhelpful in answering the question whether the pricing kernel is U-shaped. However, restricting kernels to be decreasing would result in such flat parts of the kernel.

The purpose of this paper is, therefore, to check whether the pricing kernel has increasing parts and, if so, where these are. The main contributions of this paper are as follows: First, it makes clear that the increasing parts of the kernel are not an artifact of the polynomial functional forms by estimating the kernel for functional forms, where the kernel may be U-shaped or where it is not (i.e., the piecewise linear kernel and the modified quadratic kernel, which starts at one point to be linear). These various functional forms, furthermore, allow direct testing for the increasing parts in the kernels. Second, estimating a higher order polynomial kernel than the literature reveals also that the dataset, including the Fama-French value, size, and momentum portfolios, has a clearly U-shaped kernel, which was not the case with a quadratic kernel.

Following a large part of the previously cited literature, the kernel is estimated on equity data by the generalized method of moments (GMM). Following Post et al. (2008), the pricing kernel is assumed to be constant over time. There are two main reasons for this: First, the question to be answered in this paper is if the pricing kernel is persistently U-shaped. Therefore, even if the pricing kernel is time varying, the focus is on the shape of the kernel as an average over the long run. In line with that is the usage of equity data, because the available data history is much longer for them than for option data. Second, theory does not show how the time variation of the kernel should be modeled. Taking, for example, the equilibrium model provided later in this paper, the pricing kernel depends only on the preferences of the representative agent. Given that the preferences remain roughly constant over time, the pricing kernel is also. Furthermore, it is reasonable to choose a simple econometric model, which is capable of modeling the main features but is only slightly misspecified, rather than a complicated one, where one cannot be certain that it is even more misspecified. Concerning the results of the estimations, the quadratic kernel turns out be U-shaped, in line with Dittmar (2002), Potì (2006) and Post et al. (2008). Each of them estimates the kernel on one dataset. To ensure robustness of the results, this paper estimates everything using five different datasets. Additionally, this paper quantifies the effect of the increasing parts of the kernel further: For two of these datasets, the kernel lies in the increasing region for more than one-quarter of the total observed period. In the other three datasets, the kernel lies in the increasing part for less than 2.5% of the observed periods. These results demonstrate two things: the pricing kernel is U-shaped, but the number of observations, when an increasing kernel occurs, may vary substantially.

An obvious issue with a polynomial kernel, especially a quadratic kernel, is that it almost automatically has increasing parts in it. Therefore, it is interesting to see how well a kernel does that is restricted to be decreasing. For example, Post et al. (2008) restricted the parameters of the kernel in such a way that on all observed data, the kernel is falling. With such an approach, the parameters may be strongly restricted: in my dataset the observed market returns range from -0.29 to 0.384, but 95% of the market returns are observed between -0.105 and 0.1. If a falling kernel is enforced in too large a range, the parameters of the whole kernel in that setup are massively more restricted. This becomes problematic if, as shown, only 5% of the observations can be found on approximately one-half of the restricted range.

A further step for investigating the shape of the kernel is to restrict the slope of the kernel to zero in increasing areas (as in Dittmar (2002)). Since the kernels with the restricted and the unrestricted slope are not in this way nested in each other, it is noteworthy that the nonincreasing kernel fits the data worse than the kernel with the increasing parts. However, this effect is less strong than in Dittmar (2002). Setting the slope of the increasing parts to zero is one possible way, but might it be better to restrict the slope to be smaller than another level, for example

-0.01? An innovation of this paper is to estimate the level where the slope of the kernel should be restricted: it turns out that this level, in four out of five datasets, is (insignificantly) positive. That is, even if the functional form of the kernel explicitly allows for the increasing parts of the kernel to be flat, the estimation shows increasing parts.

A potential issue is the order of the polynomial; the literature, for example Dittmar (2002) and Potì (2006), stops with pricing kernels of order 3. This paper tests polynomials up to order 7. In four out of our five datasets, this is enough. However, in the fifth dataset, the Fama-French value, size, and momentum portfolios, a pricing kernel with a polynomial of at least order 6 with a clear U-shape is appropriate. Chung et al. (2006) and Nguyen and Puri (2009) showed that Fama-French value, size, and momentum excess returns can be explained by a higher-order polynomial of the market excess return; the resulting pricing kernel is, therefore, a consequence of this. Because a polynomial of order 3 is sufficient, if the momentum portfolios are not included, it can be concluded that this effect can be attributed to the momentum portfolios.

Finally, a polynomial may not be the correct functional form of the pricing kernel. To check this, a piecewise linear kernel is estimated. It turns out that the estimated piecewise linear pricing kernel has approximately the same shape as the polynomial kernel, including its increasing parts.

Overall, this paper shows, with an extremely broad range of different tests, that the pricing kernels have a U-shape and that this shape is neither the result of a misspecified functional form nor a statistical artifact. Nonetheless, one word of caution: the estimation of the kernel cannot be done very precisely. The observed substantial variation between the kernels estimated from the different datasets and the poor significance of most statistical tests demonstrate this. Nonetheless, estimations on five different datasets and over a time horizon of more than 80 years point to a U-shape for the kernel. Hence one can be confident that the increasing parts of the kernel exist and that they, therefore, should be taken into account for asset pricing.

The remainder of the paper is arranged as follows. Section 1.2 provides the model framework and section 1.3 the estimation methodology.

Section 1.4 describes the datasets, and the pricing kernels are estimated and tested in section 1.5. The effects of these kernels on the utility function of a rational, representative investor are shown in section 1.6. Finally, section 1.7 presents the conclusions.

1.2 Model framework

No arbitrage implies the existence of some risk-neutral measure π, such that for all assets k, the expected return under π is the risk-free rate R^f (Harrison and Kreps (1979))—that is,

$$R^f = \mathbb{E}_\pi(R^k) \qquad\qquad \text{for all assets } k.$$

Let p_s denote the physical probability of state $s \in \{1, \ldots, S\}$. Then,

$$R^f = \mathbb{E}_\pi(R^k) = \sum_{s=1}^{S} p_s \frac{\pi_s}{p_s} R_s^k = \sum_{s=1}^{S} p_s L_s R_s^k = \mathbb{E}_P(LR^k), \tag{1.1}$$

where $L_s = \frac{\pi_s}{p_s}$ is the pricing kernel.[2] To keep notation simple, the physical probability $\mathbb{E}_P$ is, in the future, written as $\mathbb{E}$. As Equation (1.1) holds for any assets k and j, one has

$$0 = \mathbb{E}\left(L \cdot (R^k - R^j)\right), \tag{1.2}$$

and furthermore, because π and p are probabilities and $L \geq 0$,

$$\mathbb{E}(L) = 1. \tag{1.3}$$

No arbitrage implies these conditions for the pricing kernel. Instead of an unconditional expected value, a conditional expected value—that is, $0 = \mathbb{E}\left(L(R^k - R^j)|\Omega_t\right)$, as in Dittmar (2002) and Potì (2006), for example—can be used, where Ω_t is the information available in period t. More precisely, Ω_t are the realizations of some random variables in t, on

[2] Often—for example, in Cochrane (2001)—the stochastic discount factor is used instead of the pricing kernel. Both measures for the state prices differ in a (multiplicative) constant.

which the conditional expected value is defined. How the conditioning variables included in Ω_t should be chosen is unclear. Cochrane (2001, p. 145) summarized the issue as follows: "The situation is not repaired by simple inclusion of some conditioning information. Models such as the CAPM imply a conditional linear factor model with respect to investors' information sets. However, the best we can hope to do is to test implications conditioned down on variables that we can observe and include in a test. Thus, a conditional linear factor model is not testable." A possible way to circumvent this is to take the expected value of the conditional expected value

$$0 = \mathbb{E}\left(\mathbb{E}\left(L \cdot (R^k - R^j)|\Omega_t\right)\right) = \mathbb{E}\left(L \cdot (R^k - R^j)\right).$$

That is, the conditional model implies the unconditional one if the pricing kernel, L, is constant over time. Thus, it can be concluded that the unconditional model must hold, in any case, in the long run. While this method is probably not the most efficient way to estimate a pricing kernel, it relies on only a few assumptions, and the results cannot be influenced by wrongly chosen conditional variables.

1.2.1 Relation to the utility function of a representative investor

No arbitrage implies the existence of a positive pricing kernel. To obtain more information on the shape of the pricing kernel and on which variables the kernel depends, an equilibrium model can be used. Consider a two-period model in which a representative agent has initial wealth w_0 and an increasing, strictly concave, and differentiable utility function. Then the representative agent maximizes his utility:

$$\max_{\lambda_k} u(c_0) + \delta\mathbb{E}(u(C)),$$

where δ is the time discount. The maximization is done under the following budget constraints:

$$C = w_0 \left(\sum_{k=1}^{K} \lambda_k R^k + \left(1 - \sum_{k=1}^{K} \lambda_k \right) R_0 \right)$$
$$c_0 = w_0 \left(1 - \sum_{k=1}^{K} \lambda_k \right),$$

where c_0 is consumption in the first period, as given by the initial wealth minus the investment, into a portfolio of assets where λ_k is the portfolio weight of asset $k \in \{1, \ldots, K\}$, and C_s is consumption in state s, given by the payoffs of the portfolio bought in the first period. Then the budget constraints are inserted into the maximization problem. The first-order conditions of the utility maximization problem then imply for all assets k

$$R^f = \mathbb{E} \left(\delta \frac{u'(C)}{u'(c_0)} R^k \right).$$

A comparison with equation (1.1) immediately shows that the pricing kernel is given by

$$L = \delta \frac{u'(C)}{u'(c_0)}. \tag{1.4}$$

In the previous model, δ and c_0 are constants. The likelihood ratio process is, therefore, proportional to the marginal utility in the next period. Further, it is falling in consumption because of the concave utility function; that is, the increasing parts in the likelihood ratio process cannot be explained by this model.

The estimated pricing kernel supports the model as long as the pricing kernel is nonincreasing. However, empirical evidence shows that there may be increasing parts in the pricing kernel (see, for example, Aït-Sahalia and Lo (2000) for evidence from option data or Dittmar (2002) for evidence based on equity data). In this situation, any assumption of the model may be violated. For instance, Ziegler (2007) identifies problems of aggregation, misestimated beliefs of the agents, Peso problems, and heterogeneous beliefs as possible reasons for the observed increasing parts in the pricing kernel. However, several examples and some empirical evidence show that none of these explanations may suffice to explain the observed increasing parts of the kernel under a reasonable set of assumptions. Hens and Reichlin (2010) show, furthermore, with simple examples that a nonconcave utility function of the representative investor, incomplete markets, or heterogeneous beliefs may explain the increasing parts in the kernel. In this setup, the latest solution seems to be the best explanation since the other two possibilities need unrealistic assumptions or the results are fragile if the parameters are changed a bit. Over all, it is challenging to explain the increasing parts of the kernel theoretically under plausible assumptions.

With the equilibrium argument above, the pricing kernel depends on the consumption of the representative investor. This consumption is typically approximated by the use of either aggregate consumption or market portfolio return data. The latter source assumes that the only source of income for an investor is his assets and that by market clearing, the representative investor holds all assets in equilibrium and, therefore, earns the market return. Mankiw and Shapiro (1986) have shown that using market portfolio returns as a consumption proxy explains the observed risk premia (in a CAPM setup) much better than does using aggregated consumption. The main competitor for market portfolio returns—namely, aggregate consumption data from the National Income and Products Accounts, which has been used by Hansen and Singleton (1982) and many others—causes several problems. For example, Breeden et al. (1989), Ferson and Harvey (1992), Wilcox (1992), and Slesnick (1998) discuss measurement errors, definitional problems, issues with seasonal adjustment, and other problems with the aggregation of the consumption data over time. Furthermore, many behavioral explanations—for example, narrow framing, loss aversion, or mental accounting—demonstrate why wealth (or changes in it) should be included in the utility function of the representative agent and also, therefore, in the pricing kernel. St-Amour (2007) gives a short overview of this literature. For example, Barberis and Huang (2006) and Hens and Wöhrmann (2006) demonstrated that the equity premium puzzle can be explained in that way. For these reasons, the pricing kernel is chosen as a function of the market excess return: $L(r_m)$ with $r_m = R^M - R^f$. A side effect of this is that by subtracting the riskless rate, the kernel is in real terms. This makes sense because the consumer is not interested in his nominal wealth but in the amount of real consumption he can afford with his wealth. Nonetheless, the impact of stochastic inflation on portfolio decisions may be complex, as demonstrated by Brennan and Xia (2002) in a continuous time setup. The chosen approach tries, therefore, to keep the impact of stochastic inflation on our results as small as possible but does not claim to solve this issue.

1.2.2 Functional form of the pricing kernel

To test for increasing parts, kernels that can, but do not have to, contain increasing parts are especially interesting. In the following, we consider linear, polynomial, and piecewise linear kernels. The first two types are important since they are used in a large part of the literature. However, many polynomial, and especially quadratic, kernels have the disadvantage that almost by definition they have increasing parts. Because of that, the piecewise linear kernels are also used.

In the CAPM, the pricing kernel is a linear function of market excess return. Rubinstein (1973) showed that the CAPM can be considered as a first-order Taylor approximation of a representative investor with an arbitrary utility function. A higher-order Taylor approximation of the pricing kernel $L = \delta u'(C)/u'(c_0) = \delta u'(r_m)/u'(c_0)$ at the risk-free rate (assuming that c_0 is constant) has the following form:

$$L(r_m) = h_0 + h_1 u'' \cdot r_m + h_2 u''' \cdot r_m^2 + h_3 u'''' \cdot r_m^3 + \dots$$

The quadratic and cubic terms can be interpreted as preferences for skewness and kurtosis. Typically, most people prefer positively skewed distributions without fat tails; therefore, $h_2 u'''$ should be positive, and $h_3 u''''$ should be negative. The polynomial kernel is defined as

$$L(r_m) = \theta_0 + \theta_1 \cdot r_m + \theta_2 \cdot r_m^2 + \theta_3 \cdot r_m^3 + \dots = \theta_0 + \sum_j \theta_i \cdot r_m^j. \quad (1.5)$$

This kernel has two advantages: it is extremely general (every continuous function can be approximated by it) and it can be written in terms of linear factors, where $r_m, r_m^2, \ldots$ are the factors. However, is a polynomial the right type of function to approximate a pricing kernel? By construction, a Taylor approximation describes a function well at the point of approximation, but worsens the further it is away from that point. To estimate the pricing kernel for large or small market returns, other functional forms could potentially work better. An alternative is to use a piecewise linear kernel with breakpoints $q_1, \ldots, q_n$, i.e., as follows:

$$L(r_m) = \tau_0 + \tau_1 r_m + \begin{cases} 0 & \text{for } r_m < q_1 \\ \tau_2(r_m - q_1) & \text{for } q_1 \leq r_m < q_2 \\ \tau_2(r_m - q_1) + \tau_3(r_m - q_2) & \text{for } q_2 \leq r_m < q_3 \\ \vdots & \vdots \\ \tau_2(r_m - q_1) + \tau_3(r_m - q_2) + & \\ \qquad \ldots + \tau_{n+1}(r_m - q_n) & \text{for } q_n \leq r_m. \end{cases} \tag{1.6}$$

The main advantage of this functional form is the enormous number of possible shapes of the kernel. Post and van Vliet (2006) is one of the few studies where a piecewise linear marginal utility function (i.e., in a representative agent model, a piecewise linear pricing kernel) has been used. The authors mainly find that the market portfolio is not mean variance efficient but that third-order stochastic dominance seems to hold for the market portfolio. Since they are focusing on a linear program to check the stochastic dominance, they focus on decreasing kernels. The main reason for the rare usage of the piecewise linear kernel may be that the first derivative in r_m is not continuous in all points. For the generalized method of moments (GMM) estimation, it is required only that the piecewise linear pricing kernel is differentiable in all τ, which is obviously given. Nevertheless, for numerical optimization the noncontinuous first derivative of r_m may be problematic. The estimation methods are discussed in more detail in the next section.

1.3 Estimation methods

In a further step, the pricing kernel must be estimated from data. This can be done in various ways: Based on no-arbitrage, this step is especially easy for a linear kernel, which can be determined out of the market portfolio and the risk-free asset, and it is then identical with the famous CAPM. If it is possible to represent the kernel in a linear form in factors, then the kernel can be estimated via OLS regressions. More complicated kernels may be estimated by GMM. All these estimation methods are discussed in more detail in the following sections.

1.3.1 Benchmark CAPM

The CAPM of Sharpe (1964), Lintner (1965) and Mossin (1966) assumes a linear pricing kernel $L = \tilde{\theta}_0 + \tilde{\theta}_1 r_m$ with market excess return $r_m = R^M - R^f$. This pricing kernel can easily be estimated out of the return of the market portfolio and the riskless asset. Assuming that asset k in equation (1.2) is the market portfolio and that asset j is the risk-free rate, the following conditions are satisfied:

$$\begin{aligned} 0 &= \mathbb{E}\left(L \cdot (R^M - R^f)\right) = \tilde{\theta}_0 \mathbb{E}(r_m) + \tilde{\theta}_1 \mathbb{E}(r_m^2) \\ 1 &= \mathbb{E}(L) = \tilde{\theta}_0 + \tilde{\theta}_1 \mathbb{E}(r_m). \end{aligned}$$

Solving this system of two linear equations for $\tilde{\theta}_0$ and $\tilde{\theta}_1$ results in the following:

$$\tilde{\theta}_0 = \frac{\mathbb{E}(r_m^2)}{\mathrm{var}(r_m)} \qquad \text{and} \qquad \tilde{\theta}_1 = -\frac{\mathbb{E}(r_m)}{\mathrm{var}(r_m)}. \tag{1.7}$$

Plugging in the average and variance of past market portfolio excess returns is the simplest way to estimate the parameters of a linear pricing kernel. The advantage of this method of estimating a pricing kernel is that it depends only on the risk-free rate and the return of the market portfolio. Therefore, it does not depend on the returns of individual assets. This is an advantage, because it is typically not feasible to include every single asset in the world in an empirical study; therefore, the choice of the assets to include may affect the results. For the rest of the paper, this estimation method for a pricing kernel is referred to as the benchmark CAPM model. The following two subsections describe two additional methods for estimating more general pricing kernels.

1.3.2 Factor models

If pricing kernels are linear combinations of factors, they can be estimated using linear factor models. The CAPM (or its higher moment versions) or the Fama-French three-factor model are special cases of factor models. Furthermore, all the functional forms of the previous section can be rewritten in factor form. Since a large part of the literature focuses on factor models and on the risk premia for the different factors, it is helpful to illustrate their link with the pricing kernel. Assume that the pricing kernel is given by $L = b_0 + \boldsymbol{b}'\boldsymbol{f}$, where $\boldsymbol{f}$ is a vector of factors that vary over time, and b_0 and the vector b are constants. Given $\mathbb{E}(L) = 1$, this can be rewritten as $L = 1 + \boldsymbol{b}'\left(\boldsymbol{f} - \mathbb{E}(\boldsymbol{f})\right)$. Writing everything in terms of excess returns, $R^e_k = R^k - R^f$, Equation (1.2) becomes

$$0 = \mathbb{E}\left(L \cdot (R^k - R^f)\right) =: \mathbb{E}\left(L \cdot R^e_k\right).$$

With $\boldsymbol{R^e}$ defined as the vector of excess returns of all assets, this is equivalent to

$$\begin{aligned}
0 &= \mathbb{E}\left(L \cdot \boldsymbol{R^e}\right) = \mathbb{E}(\boldsymbol{R^e}) + \boldsymbol{b}' \operatorname{cov}\left(\boldsymbol{f}, \boldsymbol{R^e}\right) \\
\mathbb{E}(\boldsymbol{R^e}) &= -\boldsymbol{b}' \operatorname{cov}\left(\boldsymbol{f}, \boldsymbol{R^e}\right) = -\boldsymbol{b}' \operatorname{var}(\boldsymbol{f}) \operatorname{var}(\boldsymbol{f})^{-1} \operatorname{cov}(\boldsymbol{f}, \boldsymbol{R^e}) \\
&= \boldsymbol{\lambda}'\boldsymbol{\beta},
\end{aligned}$$

where

$$\boldsymbol{\lambda} = -\operatorname{var}(\boldsymbol{f})\boldsymbol{b} \qquad \text{and} \qquad \boldsymbol{\beta} = \operatorname{var}(\boldsymbol{f})^{-1} \operatorname{cov}(\boldsymbol{f}, \boldsymbol{R^e})$$

are the risk premium and risk exposure of every risk factor. The CAPM is a special case of that model. If the CAPM pricing kernel $L = \tilde{\theta}_0 + \tilde{\theta}_1 r_m$ is plugged in, the expected excess return of an asset in the CAPM is a risk premium, $\tilde{\lambda} = \mathbb{E}(r_m)$,[3] times the CAPM-$\beta$. It measures the exposure to the market risk and is defined by $\tilde{\beta} = \operatorname{cov}(r_m, R^e)/\operatorname{var}(r_m)$.

In general, $\boldsymbol{\beta}$ are the estimated multiple regression coefficients of excess returns R^e on the demeaned factors $\boldsymbol{f}$. This offers one way to estimate $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$: first, regress $\boldsymbol{R^e}$ on the demeaned factors $\boldsymbol{f}$ to obtain the estimator $\hat{\beta}$. Second, regress the average excess returns of the assets $\boldsymbol{\bar{R}^e}$ on $\boldsymbol{\hat{\beta}}$ to obtain an estimator for the risk premium $\boldsymbol{\hat{\lambda}}$. Most of the extant literature stops at this point and estimates the risk premia, λ_k, of each factor. Because $\boldsymbol{b} = -\operatorname{var}(\boldsymbol{f})^{-1}\boldsymbol{\lambda}$, the estimator for $\boldsymbol{b}$ is

$$\hat{\boldsymbol{b}} = -\left(\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T} f_t\right)\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)'\right)^{-1}\boldsymbol{\hat{\lambda}}.$$

The estimated pricing kernel, $\hat{L}$, is obtained by plugging $\hat{\boldsymbol{b}}$ into the definition of the pricing kernel:

$$\hat{L} = 1 + \hat{\boldsymbol{b}}'\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right).$$

Appendix 1.8 shows that under weak assumptions (heteroskedasticity is allowed), $\hat{\boldsymbol{b}}$ is a consistent estimator of $\boldsymbol{b}$. Through the consistency of $\hat{\boldsymbol{b}}$ and the law of large numbers, it directly follows that $\hat{L}$ is consistent. While consistency alone does not provide any information about confidence intervals or the significance of the parameters, these can be obtained using GMM estimation.

1.3.3 Estimation via the generalized method of moments

Compared to the factor model, GMM permits more general pricing kernels and provides asymptotic test statistics for the estimated parameters. Assuming that the pricing kernel is a function of the market excess return

[3] To obtain that, plug equation (1.7) into the definition of λ.

r_m, equations (1.2) and (1.3) imply the following moment restrictions:

$$0 = \mathbb{E}\left(L(r_m) \cdot (R^k - R^f)\right)$$
$$0 = \mathbb{E}\left(L(r_m)\right) - 1.$$

The market portfolio also has to be priced correctly. Therefore,

$$0 = \mathbb{E}\left(L(r_m) \cdot r_m\right) = \sum_k \mathbb{E}\left(L(r_m) \cdot \omega_k (R^k - R^f)\right),$$

where ω_k is the weight of asset k in the market portfolio in every time period. Moreover, this condition ensures that the sum of the pricing error over all assets is close to zero. Given these moment conditions, the parameter of the pricing kernel can be estimated using GMM. Note that no assumptions about the distribution of the returns are required, only that all moments must exist. Additionally, some regularity conditions should be satisfied.

Assume that $\boldsymbol{X}_t$ is a vector of the data needed to estimate the model (mainly asset and market returns in t); then $\boldsymbol{\theta_T}$ is the vector of the true parameter of the pricing kernel, and the vector of all moment conditions is written as

$$0 = \mathbb{E}(\boldsymbol{g}(\boldsymbol{X}_t, \boldsymbol{\theta_T})) := \mathbb{E}\begin{pmatrix} L(r_m, \boldsymbol{\theta_T}) \cdot (R^1 - R^f) \\ \vdots \\ L(r_m, \boldsymbol{\theta_T}) \cdot (R^K - R^f) \\ \sum_k L(r_m, \boldsymbol{\theta_T}) \cdot \omega_k (R^k - R^f) \\ L(r_m, \boldsymbol{\theta_T}) - 1 \end{pmatrix}.$$

The estimated parameter $\hat{\boldsymbol{\theta}}$ is chosen such that the deviations from the moment conditions are minimized (the deviations are weighted by a weighting matrix $\boldsymbol{W}$):

$$\hat{\boldsymbol{\theta}}(\boldsymbol{W}) = \operatorname{argmin}_{\boldsymbol{\theta}}\, n \boldsymbol{g_n}(\boldsymbol{\theta})' \boldsymbol{W} \boldsymbol{g_n}(\boldsymbol{\theta})$$
$$\boldsymbol{g_n}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{X}_t, \boldsymbol{\theta}).$$

Under some regularity conditions, the estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{W})$ is consistent, and if $\boldsymbol{W}$ is a consistent estimator of the inverse covariance matrix of

$\boldsymbol{g}(\boldsymbol{X}_t, \boldsymbol{\theta})$, the estimator is asymptotically efficient. The difference between the estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{W})$ and the true parameter $\boldsymbol{\theta_T}$ is asymptotically normally distributed. This approach allows all kinds of statistical tests, especially t-tests.

The minimization problem of the GMM estimator is solved numerically. Therefore, accurate starting values are crucial. Because the factor model presented before provides a consistent estimator of the model parameters, I use them as starting values.

To obtain an estimation of $\boldsymbol{W}$, the two-step GMM method in Hansen (1982) is applied.[4] Using this method, the model is first estimated with $\boldsymbol{W}$ as the identity matrix and the parameter estimates from the factor regression as starting values. Then a heteroskedasticity and autocorrelation consistent (HAC) covariance matrix for $\boldsymbol{g}(\boldsymbol{X}_t, \boldsymbol{\theta})$ is calculated (see Newey and West (1987)). With the inverse of this covariance matrix serving as the weighting matrix $\boldsymbol{W}$, the definitive model is estimated. Hansen et al. (1996) suggested more sophisticated GMM estimators: the iterated GMM and continuously updated GMM estimator, which differ in the way they determine $\boldsymbol{W}$. Newey and Smith (2004) and Anatolyev (2005) concluded that the two-step and iterated estimators are asymptotically equivalent and that the continuously updated estimator has a smaller asymptotic second-order bias than the other two estimators. With finite samples, these results can obviously differ. In this paper, the two-step estimator is used because it turned out to be the numerically most robust estimator. The iterated and continuously updated estimator leads to extremely volatile pricing kernels. (Chapman (1997) found similar problems with the iterated GMM estimator, in which the observed average returns are far from the mean returns predicted by the model.)

	17 Industry Portfolios									
Industry	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
Food	1000	0.0069	0.049	-0.02	9.4	0.8	2.2	0.8	0.09	0.005
Mines	1000	0.0071	0.068	-0.12	5.3	0.9	1.1	0.7	0.07	0.028
Oil	1000	0.0078	0.061	0.29	7.0	0.9	3.1	0.9	0.01	0.843
Clths	1000	0.0057	0.062	0.33	8.2	0.9	2.5	0.8	0.17	0.000
Durbl	1000	0.0060	0.078	1.36	19.7	1.3	7.3	1.5	0.21	0.000
Chems	1000	0.0074	0.064	0.34	9.5	1.0	4.2	1.1	0.11	0.001
Cnsum	1000	0.0072	0.050	0.25	9.2	0.7	2.8	0.8	0.06	0.070
Cnstr	1000	0.0064	0.069	0.43	8.8	1.2	4.7	1.2	0.11	0.001
Steel	1000	0.0065	0.086	1.37	16.7	1.4	9.1	1.6	0.11	0.000
FabPr	1000	0.0060	0.061	0.15	9.1	1.0	3.6	1.0	0.10	0.001
Machn	1000	0.0078	0.072	0.18	8.6	1.2	4.6	1.2	0.11	0.000
Cars	1000	0.0078	0.079	1.15	16.7	1.2	6.7	1.5	0.15	0.000
Trans	1000	0.0063	0.072	1.01	15.2	1.2	6.1	1.3	0.15	0.000
Utils	1000	0.0056	0.057	0.13	10.5	0.8	2.8	0.9	0.11	0.000
Rtail	1000	0.0068	0.060	-0.01	8.1	0.9	2.7	1.0	0.14	0.000
Finan	1000	0.0071	0.070	0.56	14.3	1.2	5.1	1.3	0.17	0.000
Other	1000	0.0056	0.052	-0.19	6.8	0.9	1.8	0.8	0.12	0.000
Market	1000	0.0091	0.055	0.14	10.5	1.0	3.5	1.0	0.12	0.000

Table 1.1: Summary statistics of the monthly excess returns (i.e. $R^k - R^f$) of the 17 industry portfolios and the market portfolio: average; standard deviations; skewness; kurtosis; CAPM-β; γ, the exposure to the skewness risk (see equation 1.8); δ, the exposure to the kurtosis risk (see equation 1.9) and the first-order autocorrelation coefficient (with p-value) of portfolio excess returns.

1.4 Data

The suggested estimation methods need the excess returns (i.e., $R^k - R^f$) of various assets and the market, and the market capitalizations of all the assets as well. Monthly data are used, starting in 1926, from the Center for Research in Security Prices (CRSP) at the University of Chicago. The CRSP all-share index, a value-weighted index of all common stocks listed on the NYSE, AMEX, and NASDAQ markets, is taken as a proxy for the market portfolio. The risk-free rate is the one-month T-Bill rate. To avoid spurious results, all kernels are estimated using the monthly excess returns of five different sets of data. The first two sets of assets are the 17 and 30 Fama-French industry portfolios.[5] They group all NYSE, AMEX, and NASDAQ stocks into 17 (30) industry

[4] All estimations are done with R. For the GMM estimations, the GMM-package of Chaussé (2010) has been used.

[5] Fama and French used data from CRSP to calculate their returns. Compustat data are used for the portfolio weights of the value portfolios.

					30 Industry Portfolios					
Industry	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
Food	1000	0.0068	0.049	0.05	9.4	0.7	2.3	0.8	0.08	0.007
Beer	1000	0.0093	0.075	1.84	25.2	1.0	3.0	1.0	0.09	0.005
Smoke	1000	0.0084	0.059	0.07	6.4	0.6	2.4	0.7	0.07	0.035
Games	1000	0.0076	0.091	0.64	12.3	1.4	6.3	1.6	0.19	0.000
Books	1000	0.0060	0.071	0.52	9.7	1.1	5.0	1.2	0.18	0.000
Hshld	1000	0.0063	0.061	0.37	15.5	0.9	3.7	1.1	0.08	0.008
Clths	1000	0.0055	0.061	0.30	7.9	0.8	1.9	0.6	0.15	0.000
Hlth	1000	0.0077	0.058	0.18	10.1	0.9	3.4	0.9	0.08	0.018
Chems	1000	0.0073	0.064	0.37	9.7	1.0	4.3	1.1	0.10	0.001
Txtls	1000	0.0063	0.081	1.05	12.6	1.2	6.3	1.3	0.18	0.000
Cnstr	1000	0.0061	0.070	0.36	8.9	1.2	4.3	1.1	0.13	0.000
Steel	1000	0.0065	0.085	1.37	16.7	1.4	9.1	1.6	0.11	0.000
FabPr	1000	0.0073	0.073	0.48	10.4	1.2	5.4	1.3	0.13	0.000
ElcEq	1000	0.0088	0.078	0.60	11.6	1.3	6.1	1.4	0.10	0.001
Autos	1000	0.0076	0.081	1.23	17.4	1.2	6.9	1.5	0.15	0.000
Carry	1000	0.0080	0.078	0.49	8.4	1.2	5.0	1.2	0.11	0.001
Mines	1000	0.0067	0.073	0.13	6.7	0.9	3.1	1.0	0.05	0.108
Coal	1000	0.0098	0.092	0.87	9.8	0.8	0.1	0.3	0.04	0.217
Oil	1000	0.0077	0.061	0.29	7.0	0.9	3.1	0.9	0.01	0.829
Util	1000	0.0056	0.057	0.13	10.5	0.8	2.8	0.9	0.11	0.000
Telcm	1000	0.0051	0.046	0.00	6.2	0.7	1.2	0.6	0.09	0.007
Servs	1000	0.0089	0.086	1.11	19.2	0.8	-1.6	0.4	0.03	0.361
BusEq	1000	0.0080	0.069	-0.22	6.1	1.1	2.1	0.9	0.09	0.005
Paper	1000	0.0071	0.061	0.36	9.5	1.0	3.9	1.0	0.06	0.071
Trans	1000	0.0059	0.073	1.10	16.0	1.1	6.3	1.3	0.15	0.000
Whlsl	1000	0.0052	0.075	0.65	14.3	1.1	4.2	1.2	0.19	0.000
Rtail	1000	0.0069	0.060	0.02	8.0	0.9	2.8	1.0	0.14	0.000
Meals	1000	0.0073	0.067	-0.34	5.6	1.0	-0.4	0.7	0.15	0.000
Fin	1000	0.0071	0.070	0.56	14.3	1.2	5.1	1.3	0.17	0.000
Other	1000	0.0048	0.069	0.36	9.1	1.1	4.3	1.1	0.14	0.000

Table 1.2: Summary statistics of the monthly excess returns (i.e. $R^k - R^f$)) of the 30 industry portfolios: average; standard deviations; skewness; kurtosis; CAPM-β; γ, the exposure to the skewness risk (see equation 1.8); δ, the exposure to the kurtosis risk (see equation 1.9) and the first-order autocorrelation coefficient (with p-value) of portfolio excess returns.

Panel A: Value Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=low	1000	0.0054	0.058	-0.02	7.9	1.0	3.0	1.0	0.13	0.000
2	1000	0.0064	0.055	-0.09	8.0	1.0	2.7	0.9	0.09	0.003
3	1000	0.0063	0.054	-0.22	7.8	0.9	2.3	0.9	0.07	0.039
4	1000	0.0062	0.061	1.26	18.8	1.1	6.4	1.3	0.17	0.000
5	1000	0.0069	0.057	0.85	15.3	1.0	5.1	1.1	0.14	0.000
6	1000	0.0073	0.062	0.95	19.2	1.1	5.6	1.3	0.17	0.000
7	1000	0.0074	0.067	1.84	23.4	1.1	8.1	1.4	0.16	0.000
8	1000	0.0090	0.070	2.13	27.2	1.2	8.9	1.5	0.19	0.000
9	1000	0.0098	0.076	1.33	17.3	1.2	7.5	1.5	0.14	0.000
10=high	1000	0.0106	0.094	2.41	27.3	1.5	11.1	1.9	0.16	0.000

Panel B: Size Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=small	1000	0.0115	0.103	3.71	39.6	1.4	11.9	1.8	0.22	0.000
2	1000	0.0096	0.090	2.27	25.0	1.4	8.5	1.6	0.19	0.000
3	1000	0.0095	0.082	1.94	23.3	1.3	8.1	1.5	0.22	0.000
4	1000	0.0090	0.076	1.56	18.8	1.3	7.3	1.4	0.19	0.000
5	1000	0.0086	0.073	1.16	16.1	1.2	6.6	1.4	0.18	0.000
6	1000	0.0085	0.070	1.04	15.1	1.2	6.5	1.4	0.18	0.000
7	1000	0.0081	0.066	0.81	14.0	1.2	5.7	1.3	0.16	0.000
8	1000	0.0074	0.062	0.76	13.8	1.1	5.4	1.2	0.14	0.000
9	1000	0.0069	0.059	0.57	13.4	1.1	4.8	1.2	0.12	0.000
10=big	1000	0.0056	0.051	0.09	9.4	0.9	3.1	0.9	0.09	0.006

Panel C: Momentum Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=low	995	0.0001	0.099	1.84	19.2	1.6	11.1	1.9	0.16	0.000
2	995	0.0040	0.083	1.84	23.2	1.3	8.9	1.7	0.15	0.000
3	995	0.0041	0.071	1.53	21.8	1.2	7.4	1.5	0.13	0.000
4	995	0.0055	0.065	1.55	20.5	1.1	7.3	1.4	0.13	0.000
5	995	0.0055	0.061	1.31	20.4	1.0	5.7	1.2	0.11	0.000
6	995	0.0062	0.059	0.76	14.8	1.0	4.9	1.2	0.11	0.001
7	995	0.0070	0.056	0.18	10.4	1.0	3.5	1.0	0.07	0.036
8	995	0.0082	0.054	0.04	7.7	0.9	2.6	0.9	0.09	0.006
9	995	0.0089	0.057	-0.30	6.6	1.0	1.5	0.8	0.06	0.061
10=high	995	0.0121	0.066	-0.51	5.2	1.0	0.1	0.7	0.08	0.016

Table 1.3: Summary statistics of the monthly excess returns (i.e. $R^k - R^f$) of the size, value, and momentum decile portfolios: average; standard deviations; skewness; kurtosis; CAPM-β; γ, the exposure to the skewness risk (see equation 1.8); δ, the exposure to the kurtosis risk (see equation 1.9), and the first-order autocorrelation coefficient (with p-value) of portfolio excess returns.

sectors, based on the SIC codes of the previous year.[6] The advantage of industry portfolios is that while similar companies are in the same industry category, the differences between the different industries are considerable. Industry portfolios, therefore, give a broad overview of the economy. An alternative way to group firms into different portfolios is to take some criterion and then form decile portfolios. Doing so results in a large spread of the chosen criterion between the portfolios. Black et al. (1972) were the first to use this method by grouping portfolios based on the past CAPM-$\beta = \operatorname{cov}(r_m, R^e)/\operatorname{var}(r_m)$ of the assets. Later, Fama and French (1992) and many others used value and size portfolios to study size and value anomalies. However, this method may also yield spurious results because of data snooping (see Lo and MacKinlay (1990) and Conrad et al. (2003)). From the Fama-French data library, the value, size, and momentum decile portfolios are used. The 10-value portfolios are formed every July by means of sorting the book-to-market ratio of the previous year. The size decile portfolios for July until the following June are based on the market capitalization in June of the previous year from all available assets listed on the NYSE, AMEX, and NASDAQ. The momentum decile portfolios are calculated based on the returns between $t-2$ and $t-12$. To retain an asset in a momentum portfolio, the prices in $t-13$ and the capitalization of that asset in $t-1$ must be available. An extension of Black et al. (1972) is to use higher moment risk factors instead of the CAPM risk factor, β, for forming decile portfolios. The aim of this procedure is to obtain assets that have risk exposures that are as different from higher order risk as possible. Analogous to Kraus and Litzenberger (1976) and Post et al. (2008), decile portfolios based on the skewness risk

$$\gamma^i = \frac{\mathbb{E}\left[\left(R^M - \mathbb{E}(R^M)\right)^2 \left(R^i - \mathbb{E}(R^i)\right)\right]}{\mathbb{E}\left[\left(R^M - \mathbb{E}(R^M)\right)^3\right]} \tag{1.8}$$

[6]The data and a detailed description of the industry sectors can be found in the Fama-French data library at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

and additionally to them on the kurtosis risk

$$\delta^i = \frac{\mathbb{E}\left[\left(R^M - \mathbb{E}(R^M)\right)^3 \left(R^i - \mathbb{E}(R^i)\right)\right]}{\mathbb{E}\left[\left(R^M - \mathbb{E}(R^M)\right)^4\right]} \tag{1.9}$$

are formed, where γ_i and δ_i are generalizations of the CAPM-β for the higher moments risk factors. The γ and δ portfolios are formed every July based on the data of the previous 36 months. The deciles are calculated on the NYSE data. The portfolios contain all the common stocks from the NYSE, AMEX, and NASDAQ, where the returns of the previous 36 months are available.

An overview of the return characteristics can be found in Tables 1.1 to 1.4. The first column includes the number of observations for each portfolio. For every portfolio, all available data points are used, so the number of observations varies slightly. Except for the momentum portfolios, all Fama-French data range from July 1926 to December 2009. The momentum portfolios begin in January 1927, as the returns for the preceding 12 months are needed to calculate momentum. Thereturn data from γ and δ portfolios are available from July 1929 to December 2009. This shorter time horizon stems from the 36-month formation period for those portfolios.

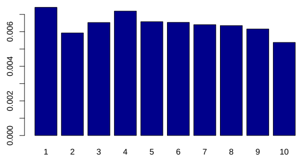
The average returns for small companies and for stocks with a low book-to-market ratio and a high past performance are better, as Figure 1.1 depicts. These effects were to be expected and were documented in such previous studies as Stattman (1980), Banz (1981), Fama and French (1992, 1993), and Jegadeesh and Titman (1993). A lower coskewness and a higher co-kurtosis also result in higher returns. The return variations between high and low γ and δ risk are considerably smaller than the value, size, and momentum effects. This may indicate that γ and δ are poor indicators for the future skewness and kurtosis risks of the assets.

For the 30 industry portfolios, the correlation between β and γ is 0.797, the correlation between β and δ is 0.899, and the correlation between γ and δ is 0.941. This means that a large part of the information on the higher moments is already in the lower moments and may indicate that incorporating higher order moments may not add much additional

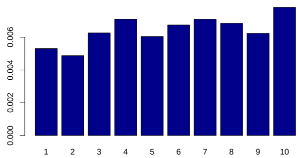
Panel A: γ-Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=low	964	0.0074	0.060	-0.20	6.4	0.9	0.8	0.8	0.06	0.070
2	964	0.0059	0.056	-0.40	8.0	0.9	1.5	0.8	0.12	0.000
3	964	0.0065	0.055	0.25	9.2	0.9	3.2	0.9	0.07	0.034
4	964	0.0072	0.057	0.76	12.1	1.0	4.5	1.1	0.10	0.002
5	964	0.0066	0.057	0.63	11.7	1.0	4.2	1.0	0.12	0.000
6	964	0.0065	0.059	0.14	7.8	1.0	3.0	1.0	0.07	0.034
7	964	0.0064	0.062	0.51	12.7	1.1	4.1	1.2	0.11	0.000
8	964	0.0064	0.069	0.80	14.8	1.2	5.4	1.3	0.14	0.000
9	964	0.0062	0.078	1.54	20.7	1.3	7.9	1.6	0.16	0.000
10=high	964	0.0054	0.087	0.99	15.4	1.5	6.7	1.6	0.14	0.000

Panel B: δ-Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=low	964	0.0053	0.043	-0.37	7.8	0.7	1.0	0.6	0.14	0.000
2	964	0.0049	0.047	-0.22	10.6	0.8	1.8	0.8	0.14	0.000
3	964	0.0063	0.050	0.22	10.0	0.8	2.9	0.9	0.07	0.022
4	964	0.0071	0.057	0.95	14.4	1.0	5.0	1.1	0.09	0.004
5	964	0.0060	0.060	0.23	9.1	1.0	3.2	1.0	0.08	0.015
6	964	0.0067	0.067	0.44	11.1	1.2	4.2	1.2	0.08	0.011
7	964	0.0071	0.071	0.70	12.0	1.2	5.4	1.3	0.07	0.026
8	964	0.0068	0.076	0.56	11.8	1.3	5.2	1.4	0.13	0.000
9	964	0.0062	0.087	0.97	14.4	1.5	7.1	1.6	0.12	0.000
10=high	964	0.0078	0.100	1.00	12.6	1.7	8.3	1.8	0.13	0.000

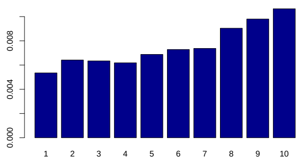
Table 1.4: Summary statistics of the monthly excess returns (i.e. $R^k - R^f$) of the γ and δ decile portfolios: Average; standard deviations; skewness; kurtosis; CAPM-β; γ, the exposure to the skewness risk (see equation 1.8); δ, the exposure to the kurtosis risk (see equation 1.9) and the first-order autocorrelation coefficient (with p-value) of portfolio excess returns.



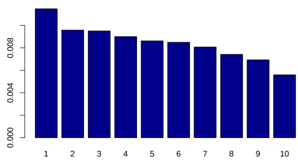
(a) Coskewness: γ



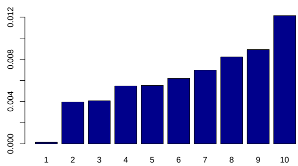
(b) Cokurtosis: δ



(c) Value



(d) Size



(e) Momentum

Figure 1.1: Average monthly excess returns (i.e. $R^k - R^f$) of the different decile portfolios

information. This high correlation is also the reason why no grouping with β-portfolios is included in the analysis: it adds no additional information.

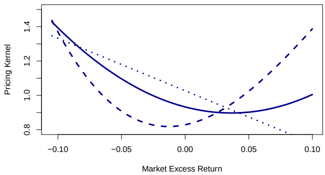
The last two columns of the tables provide the first order autocorrelation of monthly returns and the p-value for the null hypothesis that there is no autocorrelation. Significant autocorrelation can be found in the returns in a large majority of the assets. Therefore, Newey-West autocorrelation corrected standard errors will be used for all test statistics in the empirical analysis.

1.5 Empirical analysis

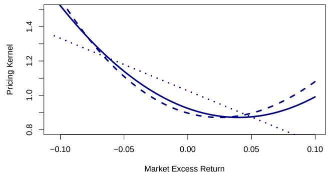
In this section, the empirical results will be discussed. In a first step, polynomial kernels up to order three are estimated. Quadratic and cubic kernels turn out to have increasing parts if estimated on industrial portfolio data. To check for the significance of the increasing parts of the kernel, the next step is to remove the increasing parts by means of a flat line. However, this makes the fit to the data poorer. A further possibility to estimate kernels is to increase the order of the polynomial further. However, except for the return data of the momentum portfolio, there is no evidence for a kernel of higher order. For the momentum portfolio data, the kernel then turns out to be clearly U-shaped. The estimation of a piecewise linear kernel and further robustness checks will confirm the previous results.

1.5.1 Linear, quadratic and cubic pricing kernels

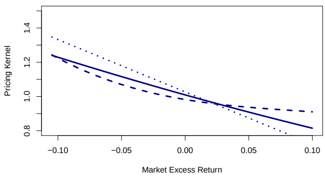
Figure 1.2 provides all of the estimated pricing kernels. The quadratic and cubic kernels are estimated by GMM. For purposes of an independent comparison, the linear benchmark CAPM kernel is also shown. The cubic and quadratic pricing kernels are similar, indicating that the cubic kernel does not behave in a totally different way from the quadratic kernel. The quadratic pricing kernels are all positive, and the shape is generally convex (for the cubic kernel evidence is more mixed), which is



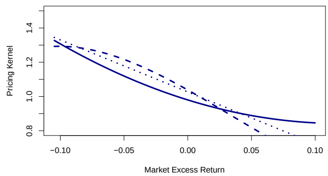
(a) 17 industries



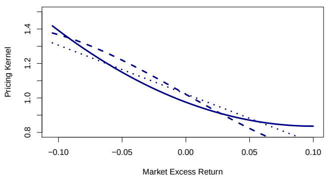
(b) 30 industries



(c) Value and size



(d) Value, size and momentum



(e) γ and δ

Figure 1.2: Estimated polynomial pricing kernels. The full line is the quadratic pricing kernel; the dashed line is the cubic pricing kernel, and the dotted line is the benchmark CAPM kernel. Estimation details for the quadratic and cubic kernel can be found in subsection 1.3.3 and in subsection 1.3.1 for the CAPM benchmark.

in line with the representative expected utility maximizer with decreasing marginal utility. Not in line with that are the increasing parts of some pricing kernels. The range of the x-axis of these figures is chosen carefully: the range between -0.105 and 0.1, which covers 95% of the market returns observed. Outside of this range, the number of observations is small; therefore, the estimation of the kernel is imprecise. A broader range on the x-axis would make the U-shapes obviously more impressive; nonetheless, only a few observations would exist in that additional area, and the kernel estimates would not be very reliable. The increasing regions observed in the industrial portfolios are because of areas where enough observations for reliable estimation are available. For the other portfolios, no evidence for increasing parts is available.

Polynomial pricing kernels up to order 3 are estimated in Table 1.5 by GMM. The J-statistic shows that a linear pricing kernel (the CAPM) is misspecified—that is, the moment conditions are statistically different from zero. Only for the dataset with the value and size portfolios does a linear pricing kernel appear to be appropriate. For the linear model with the γ and δ portfolios, which should especially take into account the higher-order risk, the J-statistic is only weakly significant at the 10% level in the linear specification. This may indicate two things: either γ and δ are poor indicators for the higher-order risk of the next year, or there is not much nonlinearity in the pricing kernel.

The J-statistic for the quadratic kernel of the momentum portfolio is still significant at the 1% level. Up to this point, everything else appears to be reasonably specified with a quadratic kernel. For example, as Kraus and Litzenberger (1976) also show, the quadratic term is positive for all portfolios. However, the quadratic parameters are not very significant. These t-statistics are in line with Potì and Wang (2010), who find the polynomial terms of order 2 or more to be insignificant. The Wald test, which tests the null hypothesis that the model is linear, shows that for the industry portfolios, a linear kernel can be rejected at the 10% significance level. The linear term in the quadratic kernel is negative, as expected from the CAPM. Moving to a cubic kernel reveals no improvement in terms of the J-statistic, and the sign of the cubic parameter is ambiguous. The Wald test also does not show large differences from a linear model. The likelihood ratio test in Table 1.10 shows that for the

	Panel A: Linear Kernel (CAPM)									
	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	1.01	**	1.02	**	1.01	**	1.01	**	1.01	**
	(0.01)		(0.01)		(0.01)		(0.01)		(0.01)	
θ_1	-2.17	**	-2.33	**	-2.20	**	-1.61	**	-2.01	**
	(0.62)		(0.60)		(0.59)		(0.57)		(0.59)	
J	33	**	51	**	20		67	**	27	

	Panel B: Quadratic Kernel									
	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	0.93	**	0.92	**	1.01	**	0.98	**	0.98	**
	(0.04)		(0.04)		(0.02)		(0.02)		(0.03)	
θ_1	-1.94	*	-2.65	**	-2.07	**	-2.31	**	-2.78	**
	(0.99)		(0.96)		(0.66)		(0.73)		(0.86)	
θ_2	26.65	†	33.26	*	1.27		9.66		13.87	
	(15.05)		(15.40)		(5.12)		(6.91)		(10.13)	
J	23		40	†	20		55	**	18	
W	3.1	†	4.7	*	0.06		2.0		1.9	

	Panel C: Cubic Kernel									
	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	0.83	**	0.90	**	0.98	**	1.04	**	1.02	**
	(0.08)		(0.07)		(0.06)		(0.04)		(0.05)	
θ_1	1.53		-2.05		-1.25		-4.18	**	-4.13	*
	(1.89)		(1.54)		(1.68)		(1.23)		(1.62)	
θ_2	54.99	*	42.53	†	8.48		-8.05		0.02	
	(27.85)		(23.95)		(18.19)		(12.20)		(17.06)	
θ_3	-143.31	†	-36.46		-32.18		82.43	†	66.85	
	(79.89)		(69.43)		(63.28)		(43.18)		(59.60)	
J	17		40	†	20		64	**	17	
W	3.9		5	†	0.26		4.4		2.7	

Table 1.5: GMM Estimation of the polynomial pricing kernels, $L = \theta_0 + \theta_1 r_m + \theta_2 r_m^2 + \theta_3 r_m^3$, for the 17 and 30 industry portfolios (Ind17 and Ind30), for the 10-value and 10-size portfolios (VS), for the 10-value, 10-size, and 10-momentum portfolios (VSM) and the γ and δ portfolios. Estimation details can be found in subsection 1.3.3. The Wald statistic tests if the quadratic (cubic) kernel is different from the linear kernel. †, *, and ** indicate significance at the 10%, 5% and 1% levels, respectively.

17 industry portfolios, the cubic model is almost significantly different from the quadratic model at the 1% level. Moreover, the parameter for the cubic term is different from zero at the 10% level. For the other portfolios, there appears to be no reason to move to a cubic kernel.

Up to the momentum portfolio with its highly significant J-statistic, all models can be reasonably well estimated by a polynomial up to order 3. Some increasing regions are found in the kernel of the industry portfolios. The next step is to examine the increasing regions in more detail.

1.5.2 A closer look at the increasing regions of the kernel

If the kernel is not just linear, a quadratic pricing kernel has an increasing region. Therefore, it is first checked whether or not that region is in an area that includes some observations of r_m. If there are no observations in that area, the increasing region is irrelevant; and if there are only a few market returns in the increasing region, then the result is most likely a statistic artifact. R_{min}, the minimum of a quadratic pricing kernel, can be determined by setting the first derivative of the kernel to zero:

$$R_{min} = \frac{-\theta_1}{2\theta_2}. \tag{1.10}$$

In the cubic case, the function can have up to one local minimum and one local maximum, which are given by:

$$R_{extrema} = \frac{\theta_2 \pm \sqrt{\theta_2^2 - 3\theta_1\theta_3}}{3\theta_3}.$$

Table 1.6 provides the minima of the quadratic kernel and the local extrema of the cubic kernel. The plausibility of increasing parts of the kernel is measured by the probability that a market return is in the increasing area; for the quadratic kernel this is, for example, $p(r_m \geq R_{min})$. This probability is measured by the number of months the kernel was in an increasing area divided by the number of all observations. For the industry portfolios, a monthly return larger than 3.6 and 4%,

Kernel	Variable	Ind17	Ind30	VS	VSM	γ and δ
Quadratic	R_{min}	0.036	0.040	0.818	0.119	0.100
	$p(r_m \geq R_{min})$	0.279	0.250	0.000	0.016	0.025
Cubic	R_{min}	0.036	0.040	0.818	0.119	0.100
	$p(r_m \geq R_{min})$	0.279	0.250	0.000	0.016	0.025
	R_{max}	0.036	0.040	0.818	0.119	0.100
	$p(r_m \geq R_{max})$	0.279	0.250	0.000	0.016	0.025

Table 1.6: Global minimum (maximum), that is, the turning points of the quadratic and cubic pricing kernels and fraction of the values of the market return that are larger than the turning points.

r_{min}	$P(r_m \geq r_{min})$	Ind17	Ind30	VS	VSM	γ and δ
0.1	0.0250	0.2859	0.2139	0.0597	0.7984	0.9977
0.2	0.0050	0.1536	0.0870	0.4042	0.5713	0.4783
0.3	0.0030	0.1224	0.0622	0.6470	0.3947	0.3476
0.4	0.0000	0.1090	0.0523	0.7845	0.3212	0.2932

Table 1.7: P-values for a Wald test of the null hypothesis that r_{min} is the global minimum of the quadratic pricing kernel. The second column shows the empirical likelihood that the market return is larger than or equal to r_{min}. The test is given for the 17 and 30 industries; the value and size; the value, size, and momentum; and the γ and δ portfolios.

respectively, is sufficient for belonging to the increasing part of the kernel. This implies that in more than 25% of all time periods, the realized pricing kernel was in the increasing region. This observation is supported by the results from the cubic kernel. Not much evidence of an increasing kernel can be found in the other portfolios. With, at most, 2.6% of all months, the quadratic pricing kernel was increasing. For the value/size portfolios, the cubic kernel implies that there are no local extrema—that is, the pricing kernel is decreasing everywhere. For the value, size and momentum, and γ and δ portfolios, the local maxima are in extremely negative returns, and the local minima in extremely positive returns. Therefore, the kernel is falling.

A next step is to check if these increasing parts are statistically significant. One way to do that is to test if the minimum of the quadratic kernel is within the observed data (or at least in an area where almost no data are observed). If this can be rejected, the pricing kernel has increasing parts in the relevant range of market portfolio returns. The null hypothesis is, therefore, that the minimum of a quadratic pricing kernel is at r_{min}. Equation (1.10) implies for the null hypothesis that

$$\theta_1 + 2r_{min}\theta_2 = 0.$$

The null hypothesis of $R_{min} = r_{min}$ is tested for r_{min} of 10, 20, 30, and 40%. In the case of $r_{min} = 0.1$, only 2.5% of all market excess returns are larger than r_{min}, and in the case of $r_{min} = 0.4$, no observed market excess return is larger. If the minimum of the pricing kernel is at one of these levels, increasing parts of the pricing kernel are in areas with (almost) no observations. Therefore, they would be irrelevant; that is, if the null hypothesis of the test cannot be rejected, increasing parts in the kernel cannot be significantly statistically supported. Table 1.7 includes the p-values of the Wald test. As shown, in no case is the estimated minimum of the quadratic pricing kernel different from r_{min} at the 5% level. A model with an increasing kernel is, therefore, not significantly different from one without: that is, the increasing parts are not significant.

A quadratic kernel always has an increasing part. To ensure that this part is not just an artifact from the functional form, a new kernel is used. The basic kernel has a quadratic form, but the slope of the kernel right to the minima is set to zero: that is, after the minimum, the kernel becomes a flat line, as illustrated with the dashed line in Figure 1.3 for the estimation for the 30-industry dataset. The estimated parameters of these kernels are in Panel A of Table 1.8. The parameters themselves are similar to the quadratic kernels in Table 1.5. However, the value of the J-statistic in four of the five portfolios is larger than in the quadratic case. The J-statistic is the sum of the weighted quadratic moment deviation divided by the number of time periods—that is, the criterion minimized by GMM. The smaller values indicate that the increasing parts in the pricing kernel improve the fit of the model.

The earlier approach can be generalized. A maximal level of the slope of the kernel, m, is fixed. If the slope of the estimated quadratic kernel would be larger than m, the slope is set to m. With $m = 0$, the kernel is as previously, and if $m = +\infty$, the kernel is a standard quadratic kernel. If the kernel were decreasing, as the representative agent model suggests, one would expect the kernel to continue at one point in the decreasing part of the U with a linear negative slope: that is, $m \leq 0$. $m > 0$ (given that $\theta_2 > 0$) indicates the opposite: the linear part starts after the minima and is increasing. The dotted line in Figure 1.3 illustrates this. With $m = 8.71$, the positive slope starts far outside the plotted range; therefore, the kernel is U-shaped. An estimate of m can be found in the lower part of Table 1.8. m is positive in four of the five portfolios such that the kernel contains an increasing part. Nonetheless, m is not significantly different from zero; that is, there is no statistically significant evidence for increasing parts in the pricing kernel.

1.5.3 Higher-order pricing kernel

Up to now, only polynomials up to the third order have been taken into account. Higher-order polynomials may reveal even more information. In the next step, polynomials up to order 7 are considered. Table 1.9 gives the p-values of the J-statistics that check if the moment conditions are satisfied for the different polynomials. For quadratic and cubic kernels, the J-statistic is not significant for most portfolios. Only for the value, size, and momentum portfolio does a kernel above order 3 help. In that instance, the J-statistics are highly significant until order 5 and not significant after orders 6 and 7. In the case of the 30 industry, the J-statistics turn out to be insignificant for a quadratic kernel but significant for some higher-order polynomials. Intuitively, one would expect that a higher-order kernel would always fit the data better than a lower order kernel and, therefore, that the J-statistic would fall with the order of the polynomial since a higher-order kernel is, by definition, always able to fit the data at least as well as a kernel of higher order. However, the more parameters the model has, the less over-identifying restrictions exist; therefore, the degrees of freedom of the χ^2 distribution of the J-statistic become smaller with the higher polynomial order of the kernel, and this has a decreasing effect on the p-value. The first effect is typ-

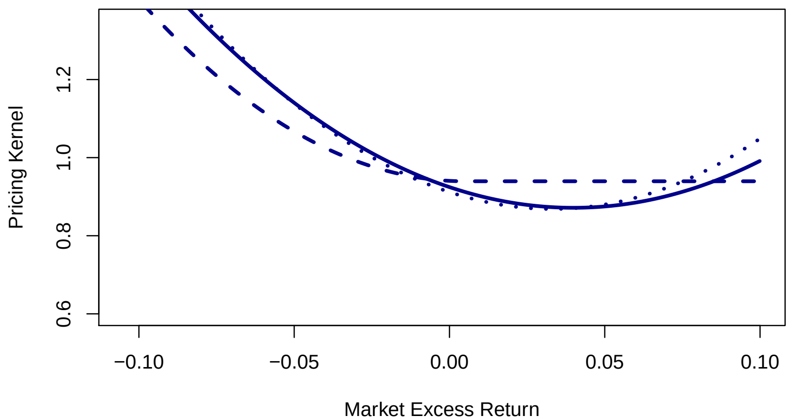


Figure 1.3: Estimated pricing kernels for the 30 industry portfolios. The full line is the quadratic pricing kernel; the dashed line is the quadratic kernel, which is flat after the minimum; and the dotted line is the quadratic kernel, which continues linear after the slope 8.71. Estimation details can be found in subsection 1.3.3.

	Panel A: Quadratic Kernel Flat after Minimum									
	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	2.4		0.94	**	1.01	**	0.96	**	0.99	**
	(5.5)		(0.02)		(0.02)		(0.05)		(0.05)	
θ_1	33.6		-0.42		-2.07	**	-1.51		-2.40	**
	(93.1)		(5.53)		(0.66)		(2.30)		(0.77)	
θ_2	194.9		42.24		1.27		19.86		6.89	
	(337.3)		(50.92)		(5.12)		(31.20)		(19.38)	
J	22		47	*	20		60	**	23	

	Panel B: Quadratic Kernel until Slope is m, afterwards Linear									
	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	0.84	**	0.91	**	1.0		1.04	**	0.97	**
	(0.13)		(0.08)		(2.2)		(0.04)		(0.04)	
θ_1	2.61		-2.61	*	-1.5		-4.18	**	-2.91	†
	(8.32)		(1.17)		(42.8)		(1.23)		(1.51)	
θ_2	90.75		38.89		3.6		-8.87		15.52	
	(68.65)		(29.54)		(154.2)		(12.11)		(13.95)	
m	3.80		8.71		-1.9		82.16	†	7.55	
	(4.37)		(7.90)		(1.6)		(42.95)		(68.62)	
J	20		40	†	20		63	**	18	

Table 1.8: Panel A shows the estimation of a quadratic pricing kernel when the kernel becomes flat and when the quadratic function is minimal. Panel B shows a quadratic kernel in which the slope is restricted to be smaller or equal to m. If the slope is in a certain range larger than m, the kernel is made linear with a slope of m in that range. Estimations were done for the 17 and 30 industry portfolios (Ind17 and Ind30), for the 10-value and 10-size portfolios (VS); for the 10-value, 10-size, and 10-momentum portfolios (VSM); and for the γ and δ portfolios. Estimation details can be found in subsection 1.3.3. The Wald statistic tests if the quadratic (cubic) kernel is different from the linear kernel. †, *, and ** indicate significance at 10%, 5%, and 1% levels, respectively.

Order	Ind17	Ind30	VS	VSM	γ and δ
1	0.010	0.009	0.466	0.000	0.143
2	0.114	0.079	0.389	0.002	0.543
3	0.342	0.065	0.358	0.000	0.490
4	0.264	0.014	0.412	0.000	0.373
5	0.249	0.028	0.364	0.000	0.405
6	0.233	0.018	0.262	0.256	0.299
7	0.469	0.044	0.237	0.275	0.678

Table 1.9: P-values of J-test: The null hypothesis H_0 is that the polynomial pricing kernels of orders 1 to 7 are able to explain the moment conditions.

ically stronger, and, therefore, the p-values are rising most of the time with the order of the polynomial of the kernel.

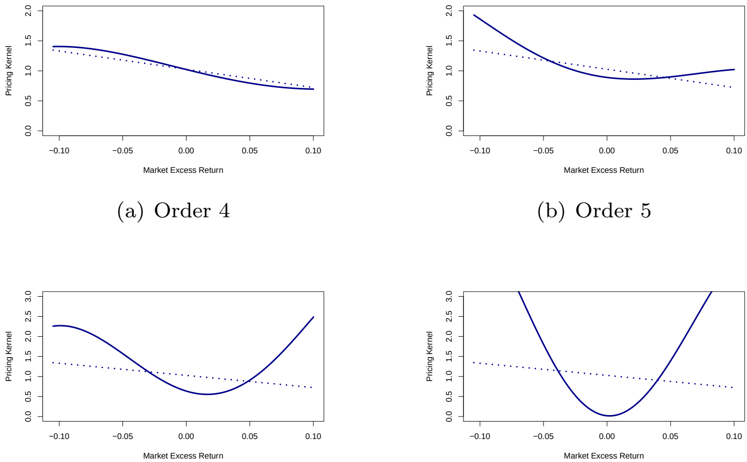
If a model with a kernel of a higher order does not perform better than one with a lower order, there is no reason to choose the model with more parameters. The likelihood ratio test checks if two nested models are statistically different. If the higher-order model is not different, then it is better to choose the lower-order model. Table 1.10 shows that the linear kernel can be rejected in almost all cases against the higher-order kernels. The exception is the value size portfolios. Linear kernels are, therefore, not sufficient. Except for the data with the momentum portfolios, all other datasets can be modeled with a cubic kernel in the case of the 17-industry dataset and a quadratic kernel for the other portfolios. In the case of the momentum portfolios, the kernels with orders 6 and 7 are always different from the kernels with orders 1 to 3. The kernels of orders 4, 5, 6, and 7 as estimated with the momentum data are shown in Figure 1.4. The kernels of orders 4 and 5 have a significant J-statistic (i.e., are misspecified) and are not significantly different from the kernels of orders 2 and 3. Put differently, they have practically no increasing parts (for that see figure 1.2(d)). The two kernels with the higher order are statistically significantly different from the lower-order kernels and are well specified. The momentum portfolio, therefore, requires a kernel of at least order 6. The momentum kernel is U-shaped and contains an increasing part. The lower-order kernels, which are unable to explain the average returns of the momentum portfolios, did not contain an

Panel A: H_0 is a Linear Kernel					
Order	Ind17	Ind30	VS	VSM	γ and δ
2	0.001	0.001	1.000	0.001	0.003
3	0.000	0.004	0.859	0.229	0.010
4	0.001	0.123	0.528	0.067	0.037
5	0.002	0.044	0.641	0.193	0.040
6	0.003	0.100	0.870	0.000	0.095
7	0.001	0.026	0.868	0.000	0.016

Panel B: H_0 is a Quadratic Kernel					
Order	Ind17	Ind30	VS	VSM	γ and δ
3	0.012	0.648	0.470	1.000	0.646
4	0.047	1.000	0.295	1.000	1.000
5	0.073	1.000	0.433	1.000	0.804
6	0.098	1.000	0.724	0.000	0.987
7	0.031	0.648	0.743	0.000	0.253

Panel C: H_0 is a Cubic Kernel					
Order	Ind17	Ind30	VS	VSM	γ and δ
4	1.000	1.000	0.166	0.040	1.000
5	0.733	1.000	0.330	0.208	0.677
6	0.689	1.000	0.673	0.000	0.987
7	0.204	0.537	0.700	0.000	0.173

Table 1.10: P-values of likelihood ratio tests. The linear, quadratic, and cubic kernels are tested against pricing kernels up to order 7. The null hypothesis H_0 is the linear, quadratic, or cubic pricing kernel.



(a) Order 4 (b) Order 5

(c) Order 6 (d) Order 7

Figure 1.4: Pricing kernels for the value, size, and momentum portfolios for polynomial kernels of orders 4 to 7. The dotted line is the benchmark CAPM. Estimation details can be found in subsection 1.3.3.

increasing part, showing also that a U-shaped kernel is needed to explain the risk premium on the momentum portfolios.

In line with Dittmar (2002) and Potì (2006), a kernel up to order 3 is required in most cases, considering the J-statistic and the likelihood ratio test. The only exception is the dataset with the momentum portfolio. However, the fact that the U-shape of the kernel becomes massively stronger with the use of a higher order kernel even strengthens the hypothesis of U-shaped kernels.

1.5.4 Piecewise linear kernel

Up to this point, the focus has been on polynomials. A major problem, especially with the quadratic kernel, is that there are increasing parts of the kernel almost by construction. To verify that these increasing parts are not an artifact of the chosen functional form, piecewise linear kernels are estimated:

$$L = \tau_0 + \tau_1 r_m + \begin{cases} 0 & \text{for } r_m < q_1 \\ \tau_2(r_m - q_1) & \text{for } q_1 \leq r_m < q_2 \\ \tau_2(r_m - q_1) + \tau_3(r_m - q_2) & \text{for } q_2 \leq r_m < q_3. \end{cases} \tag{1.11}$$

In this setup, there must not be any increasing part in the kernel. In the following, the market portfolio returns are split into three quantiles (with 33% of the observations of r_M in each), and the kernel is estimated. The estimations can be found in Table 1.11, and the piecewise linear kernel is plotted together with the quadratic kernel in Figure 1.5. The specification tests are quite similar to those for the quadratic and cubic kernels. The only exception is the kernel estimated from the γ and δ portfolios, which start with an increasing part, fall, and then increase again. This change is strong enough that a Wald test indicates that the model is statistically different from a linear model. In the plots, all kernels are more or less moving around the quadratic kernel. Four of the five portfolios show an increasing kernel in the third quantile. If this slope is significantly positive, it can be checked using a Wald test. The p-values for the 17 and 30 industry, the VSM, and the γ and δ portfolios are

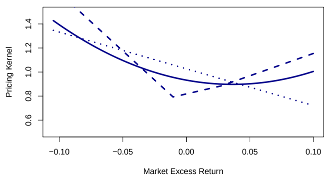
	Ind17		Ind30		VS		VSM		γ and δ	
τ_0	0.69	*	1.00	**	0.91	**	1.01	**	1.60	**
	(0.32)		(0.27)		(0.33)		(0.25)		(0.26)	
τ_1	-9.67		-5.39		-3.75		-3.80		6.39	
	(6.37)		(5.15)		(5.92)		(4.59)		(4.33)	
τ_2	12.11		-4.04		5.49		-2.93		-30.30	*
	(15.87)		(13.32)		(16.25)		(12.04)		(12.16)	
τ_3	1.28		15.52		-4.53		9.62		26.33	**
	(12.05)		(11.13)		(11.31)		(8.87)		(10.05)	
J	22		41	†	21		54	**	14	
W	2.9		4.7	†	0.2		3.5		7	*

Table 1.11: Estimation of the piecewise linear pricing kernel (i.e., equation (1.11)), for the 17 and 30 industry portfolios (Ind17 and Ind30), the 10-value and 10-size portfolios (VS), the 10-value, 10-size, and 10-momentum portfolios (VSM), and the γ and δ portfolios. Estimation details can be found in subsection 1.3.3. The Wald statistic tests whether or not the quadratic (cubic) kernel is different from the linear kernel. †, *, and ** indicate significance at 10%, 5%, and 1% levels, respectively.

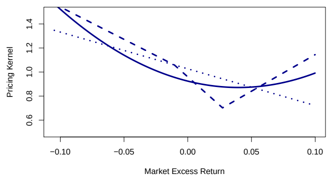
0.367, 0.142, 0.327, and 0.459. This shows again that the kernel might be U-shaped, and also that on this occasion, statistical significance is an issue.

1.5.5 Robustness checks

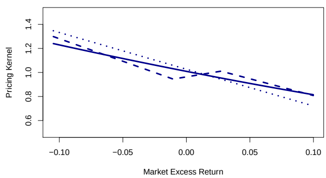
All the estimations have been made using five sets of portfolios, which can be seen as a first robustness check. A next obvious robustness check is to pool all datasets (i.e., 30 industries, value, size, momentum, and γ and δ-portfolios); the results were comparable to the results of the 30 industries-portfolio. Using more assets also did not improve the significance of the results. Further, using the identity matrix, the covariance matrix of the returns or a standard covariance matrix (not taking into account the serial correlation) as the inverse of the weighting matrix $\boldsymbol{W}$ does not change the general shape of the pricing kernels. The estimations via GMM and OLS are, furthermore, similar. The next step



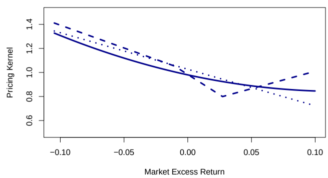
(a) 17 industries



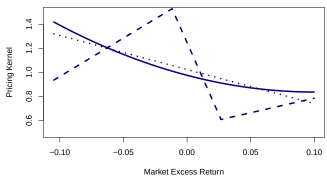
(b) 30 industries



(c) Value and size



(d) Value, size and momentum



(e) γ and δ

Figure 1.5: Piecewise liner pricing kernel (dashed line) compared with the quadratic pricing kernel (full line) and the benchmark CAPM (dotted line). Estimation details can be found in subsection 1.3.3.

Time period	Ind17	Ind30	VS	VSM	γ and δ
$\leq$ 1950		0.232			
$p(r_m \geq R_{min})$		0.014			
1951-1970	0.020	0.478		0.028	
$p(r_m \geq R_{min})$	0.417	0.000		0.346	
1971-1990		0.113	0.047	0.015	0.053
$p(r_m \geq R_{min})$		0.025	0.221	0.454	0.154
1991-2009	0.084	0.019	0.012	-0.010	
$p(r_m \geq R_{min})$	0.017	0.435	0.527	0.674	

Table 1.12: Global minimum of the quadratic pricing kernel and fraction of the values of the market return larger than the turning point. If the estimated quadratic term had a negative sign, the cell has been left empty.

is to check the time stability and to check if there are any issues with multicollinearity.

The increasing parts of the pricing kernels are the main points of interest. A good indicator for these is the global minima of a quadratic pricing kernel. In Table 1.12, these are calculated for several time windows of approximately 20 years. In cases where the quadratic function had a maximum instead of a minimum, the cells are left open. For the period before 1950, including the Great Depression and World War II, all kernels are almost linear (exact coefficients are not tabulated), and most of the second-order coefficients are slightly negative. In the case of the 30 industry portfolios, the quadratic term is slightly positive with a value of 3.64. This implies a turning point at 23.2% market returns per month. The probability that this or an even larger return occurs is 1.4%, which is extremely small. In the other much less extraordinary time periods, most pricing kernels are convex and, therefore, have a minimum. In over one-half of the sub periods after 1950, the probability of being in an increasing part of the kernel exceeds 20%. In contrast to the estimation over the whole time period in Table 1.5, the momentum portfolio already shows a quadratic kernel with increasing parts. For the whole period, a kernel of at least order 6 was needed to see this.

Panel A: Quadratic Kernel (Orthogonalized)	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	0.93	**	0.92	**	1.01	**	0.98	**	0.98	**
	(0.04)		(0.04)		(0.02)		(0.02)		(0.03)	
$\bar{\theta}_1$	-1.94	*	-2.66	**	-2.07	**	-2.31	**	-2.78	**
	(0.99)		(0.96)		(0.66)		(0.73)		(0.86)	
$\bar{\theta}_2$	26.71	†	33.33	*	1.32		9.72		13.94	
	(15.06)		(15.40)		(5.13)		(6.91)		(10.14)	
J	23		40	†	20		55	**	18	
W	3.1	†	4.7	*	0.07		2.0		1.9	

Panel B: Cubic Kernel (Orthogonalized)	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	0.83	**	0.90	**	0.98	**	1.03	**	1.02	**
	(0.08)		(0.07)		(0.06)		(0.04)		(0.05)	
$\bar{\theta}_1$	1.40		-2.11		-1.22		-4.14	**	-4.16	*
	(1.89)		(1.53)		(1.69)		(1.24)		(1.62)	
$\bar{\theta}_2$	55.57	*	41.88	†	8.82		-5.77		-0.49	
	(27.97)		(23.72)		(18.23)		(12.41)		(16.97)	
$\bar{\theta}_3$	-150.48	†	-39.74		-34.83		81.66	†	68.55	
	(84.22)		(71.80)		(66.29)		(44.86)		(61.45)	
J	17		40	†	20		65	**	18	
W	4.0		5	†	0.28		4.4		2.7	

Table 1.13: Estimation of the orthogonalized pricing kernel, $L = \theta_0 + \bar{\theta}_1 r_m + \bar{\theta}_2 \left(r_m^2 - a r_m\right) + \bar{\theta}_3 \left(r_m^3 - b r_m^2 - c r_m\right)$, kernels for the 17 and 30 industry portfolios (Ind17 and Ind30); the 10-value and the 10-size portfolios (VS); the 10-value, 10-size, and-10 momentum portfolios (VSM); and the γ and δ portfolios. Estimation details can be found in subsection 1.3.3. The Wald statistic tests whether or not the quadratic (cubic) kernel is different from the linear kernel. Estimation details can be found in subsection 1.3.3. †, *, and ** indicate significance at 10%, 5%, and 1% levels, respectively.

However, the evidence for increasing parts in the pricing kernel becomes much smaller for the industry portfolios. Overall, there is evidence for increasing pricing kernels after 1950.

Another potential issue is multicollinearity, since r_m, r_m^2, and r_m^3 are by definition correlated. To address this issue, estimations with orthogonalized regressors are performed. That is,

$$L = \theta_0 + \bar{\theta}_1 r_m + \bar{\theta}_2 \left(r_m^2 - a r_m\right) + \bar{\theta}_3 \left(r_m^3 - b r_m^2 - c r_m\right).$$

is the estimated kernel. a, b and c are defined such that

$$\begin{aligned}0 &= \operatorname{cov}(r_m, r_m^2 - ar_m) = \operatorname{cov}(r_m, r_m^3 - br_m^2 - cr_m)\\ &= \operatorname{cov}(r_m^2, r_m^3 - br_m^2 - cr_m).\end{aligned}$$

The results for the estimation of the pricing kernel with these orthogonalized factors (or polynomials) can be found in Table 1.13. As shown, these are comparable with the results in Table 1.5–that is, they show that the pricing kernel for the nonmomentum portfolios must be around orders 2 or 3. In addition, the signs of the polynomials are identical. The results are, therefore, robust for multicollinearity.

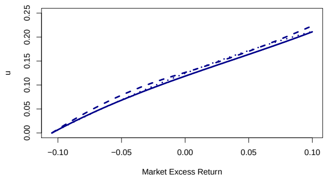
1.6 Utility function of the representative agent

Assuming there is a representative agent, what would his utility function look like? In equation (1.4), it was shown that the pricing kernel is a constant times the marginal utility, i.e. $L(r_m) = \delta \cdot u'(r_m)/u'(c_0)$. Integrating this implies

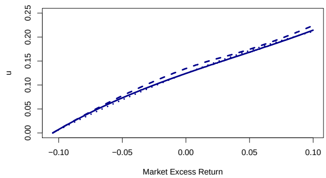
$$u(r_m) = \text{constant} + \frac{\delta}{u'(c_0)} \int_{-\infty}^{r_M} L(r) dr.$$

This analysis requires that markets be complete and that the representative agent has correct beliefs and an increasing, concave utility function. None of these assumptions must be satisfied. Nevertheless, it is interesting to see the shape of the utility function that would evolve from these assumptions.

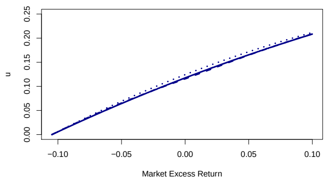
If one normalizes $\delta/u'(c_0) = 1$ and sets the utility function at the left corner of the graph to zero, the utility function implied for the quadratic, the piecewise linear, and the benchmark CAPM kernel can be found, as shown in Figure 1.6. The most obvious point is that all utility functions are quite similar. This fits the fact that it is difficult to find statistically significant differences between the different kernels. Nonconcavities are especially observed with the piecewise linear pricing kernel. The problem is that the shape of the utility function, as implied by the piecewise



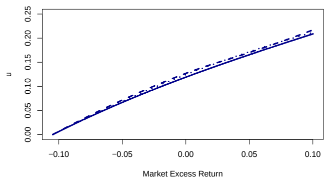
(a) 17 industries



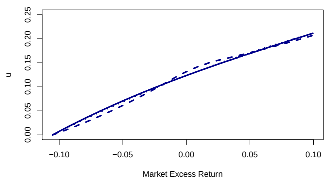
(b) 30 industries



(c) Value and size



(d) Value, size and momentum



(e) γ and δ

Figure 1.6: Utility function of the representative agent implied by the pricing kernel in the case of the quadratic pricing kernel (full line), the piecewise linear pricing kernel (dashed line), and the benchmark CAPM (dotted line).

linear kernel, is different for every portfolio. The quadratic pricing kernel is typically much closer to the CAPM benchmark. Nonconcavities are observed in the case of the industry portfolios, but even there they seem to be weak. Overall, the effect of the nonconcavities in the utility function seems to be weak. Their existence could, nonetheless, alter completely the investment behavior of the representative agent.

1.7 Conclusions

This paper examined the increasing parts of the U-shaped pricing kernels found in equity data. This has been done in a much broader way than can be found in the existing literature. In particular, the estimation on datasets in the industry and momentum portfolios shows clear evidence for increasing parts in the pricing kernel. To make sure that these increasing parts are not just an artifact of the polynomial functional form, other functional forms that allow for nonincreasing shapes lead to a poorer fit for the data. Despite the fact that the U-shape of the kernel can be shown on many different datasets, time horizons, and functional forms in terms of statistical significance, this evidence is weak. This paper shows that analogously to factor models, the value, size, and momentum effect can be explained by the polynomials of market returns of sufficiently high order. Another contribution is that the kernels of these higher-order polynomials are mainly U-shaped, increasing with positive returns. This is consistent with a positive premium on coskewness.

An implication of the increasing part of the pricing kernel is that the economy cannot be modeled by a risk-averse, utility-maximizing representative agent. This paper shows that this effect is not just a short-run phenomenon, as with the evidence from stock options data that typically holds for a specific, typically short, time period. The increasing parts in the kernel appear to persist over a time horizon of more than 80 years. So far, there is no generally accepted economic explanation for this phenomenon. For future research, it will be important to check to which degree heterogeneous, mis-estimated beliefs, Peso problems, incomplete markets, aggregation problems, and nonstandard preferences contribute to the empirically observed U-shape. To date, there is clear evidence that it is not possible to explain the whole phenomenon with only one of those factors.

Bibliography

Stanislav Anatolyev. GMM, GEL, Serial Correlation, and Asymptotic Bias. *Econometrica*, 73(3):983–1002, May 2005.

Yacine Aït-Sahalia and Andrew W. Lo. Nonparametric Risk Management and Implied Risk Aversion. *Journal of Econometrics*, 94(1-2): 9–51, January–February 2000.

Rolf W. Banz. The Relationship between Return and Market Value of Common Stocks. *Journal of Financial Economics*, 9(1):3–18, 1981.

Nicholas Barberis and Ming Huang. The Loss Aversion / Narrow Framing Approach to the Equity Premium Puzzle. NBER Working Papers 12378, National Bureau of Economic Research, Inc, July 2006.

Giovanni Barone-Adesi. Arbitrage Equilibrium with Skewed Asset Returns. *Journal of Financial and Quantitative Analysis*, 20(3):299–313, September 1985.

Giovanni Barone-Adesi and Hakim Dall'O. Is the Price Kernel Monotone? Working paper, University of Lugano, January 2010.

Giovanni Barone-Adesi, Robert F. Engle, and Loriano Mancini. A GARCH Option Pricing Model with Filtered Historical Simulation. *Review of Financial Studies*, 21(3):1223–1258, May 2008.

Fischer Black, Michael C. Jensen, and Myron S. Scholes. The Capital Asset Pricing Model: Some Empirical Tests. In Michael C. Jensen, editor, *Studies in the Theory of Capital Markets*. Praeger Publishers Inc, 1972.

Douglas T. Breeden, Michael R. Gibbons, and Robert H. Litzenberger. Empirical Test of the Consumption-Oriented CAPM. *Journal of Finance*, 44(2):231–262, June 1989.

Michael J. Brennan and Yihong Xia. Dynamic Asset Allocation under Inflation. *The Journal of Finance*, 57(3):1201–1238, June 2002.

David A. Chapman. Approximating the Asset Pricing Kernel. *Journal of Finance*, 52(4):1383–1410, September 1997.

Pierre Chaussé. Computing generalized method of moments and generalized empirical likelihood with R. *Journal of Statistical Software*, 34 (11):1–35, May 2010.

Y. Peter Chung, Herb Johnson, and Michael J. Schill. Asset Pricing When Returns Are Nonnormal: Fama French Factors versus Higher Order Systematic Comoments. *Journal of Business*, 79(2):923–940, March 2006.

John Howland Cochrane. *Asset Pricing*. Princeton University Press, 2001.

Jennifer Conrad, Michael Cooper, and Gautam Kaul. Value versus Glamour. *Journal of Finance*, 58(5):1969–1995, October 2003.

Enrico De Giorgi and Thierry Post. Second-Order Stochastic Dominance, Reward-Risk Portfolio Selection, and the CAPM. *Journal of Financial and Quantitative Analysis*, 43(02):525–546, June 2008.

Kai Detlefsen, Wolfgang Härdle, and Rouslan Moro. Empirical Pricing Kernels and Investor Preferences. SFB 649 Discussion Papers 2007-017, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany, April 2007.

Robert F. Dittmar. Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns. *Journal of Finance*, 57(1):369–403, February 2002.

Friedhelm Eicker. Limit Theorems for Regressions with Unequal and Dependent Errors. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 59–62. University of California Press, 1967.

Vihang R. Errunza and Oumar Sy. A Three-Moment International Asset-Pricing Model: Theory and Evidence. Working paper, McGill University, November 2005.

Eugene F. Fama and Kenneth R. French. The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47(2):427–465, June 1992.

Eugene F. Fama and Kenneth R. French. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1): 3–56, February 1993.

Hsing Fang and Tsong-Yue Lai. Co-Kurtosis and Capital Asset Pricing. *Financial Review*, 32(2):293–307, May 1997.

Wayne E. Ferson and Campbell R. Harvey. Seasonality and Consumption-Based Asset Pricing. *Journal of Finance*, 47(2):511–552, June 1992.

Irwin Friend and Randolph Westerfield. Co-Skewness and Capital Asset Pricing. *Journal of Finance*, 35(4):897–913, September 1980.

Yuri Golubev, Wolfgang Härdle, and Roman Timofeev. Testing Monotonicity of Pricing Kernels. SFB 649 Discussion Papers 2008-001, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany, January 2008.

James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1. edition, 1994.

Lars Peter Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–54, July 1982.

Lars Peter Hansen and Kenneth J. Singleton. Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica*, 50(5):1269–1286, September 1982.

Lars Peter Hansen and Kenneth J. Singleton. Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns. *Journal of Political Economy*, 91(2):249–265, April 1983.

Lars Peter Hansen, John Heaton, and Amir Yaron. Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business & Economic Statistics*, 14(3):262–280, July 1996.

Michael J. Harrison and David M. Kreps. Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory*, 20(3): 381–408, June 1979.

Campbell R. Harvey and Akhtar Siddique. Conditional Skewness in Asset Pricing Tests. *Journal of Finance*, 55(3):1263–1295, June 2000.

Thorsten Hens and Christian Reichlin. Three Solutions to the Pricing Kernel Puzzle. Swiss Finance Institute Research Paper 10-14, Swiss Finance Institute, March 2010.

Thorsten Hens and Peter Wöhrmann. Mental Accounting and the Equity Premium Puzzle. Working Paper 243, NCCR "Financial Valuation and Risk Management", July 2006.

Chi-Hsiou Hung. Return Predictability of Higher-Moment CAPM Market Models. *Journal of Business Finance & Accounting*, 35(7-8):998–1022, September 2008.

Jens Carsten Jackwerth. Recovering Risk Aversion from Option Prices and Realized Returns. *Review of Financial Studies*, 13(2):433–451, Summer 2000.

Narasimhan Jegadeesh and Sheridan Titman. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1):65–91, March 1993.

Alan Kraus and Robert H. Litzenberger. Skewness Preference and the Valuation of Risk Assets. *Journal of Finance*, 31(4):1085–1100, September 1976.

Kian-Guan Lim. A New Test of the Three-Moment Capital Asset Pricing Model. *Journal of Financial and Quantitative Analysis*, 24(2):205–216, June 1989.

John Lintner. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 47(1):13–37, February 1965.

Andrew W. Lo and A. Craig MacKinlay. Data-Snooping Biases in Tests of Financial Asset Pricing Models. *Review of Financial Studies*, 3(3): 431–467, Autumn 1990.

Michael Magill and Martine Quinzii. *Theory of Incomplete Markets*. MIT Press, 1996.

N. Gregory Mankiw and Matthew D. Shapiro. Risk and Return: Consumption Beta Versus Market Beta. *The Review of Economics and Statistics*, 68(3):452–459, August 1986.

Jan Mossin. Equilibrium in a Capital Asset Market. *Econometrica*, 34 (4):768–783, October 1966.

Whitney K. Newey and Richard J. Smith. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, 72(1):219–255, January 2004.

Whitney K. Newey and Kenneth D. West. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703–708, May 1987.

Duong Nguyen and Tribhuvan N. Puri. Higher-Order Systematic Comoments and Asset Pricing: New Evidence. *Financial Review*, 44(3): 345–369, August 2009.

D. F. Nicholls and A. R. Pagan. Heteroscedasticity in Models with Lagged Dependent Variables. *Econometrica*, 51(4):1233–1242, July 1983.

Thierry Post and Pim van Vliet. Downside Risk and Asset Pricing. *Journal of Banking & Finance*, 30(3):823–849, March 2006.

Thierry Post, Pim van Vliet, and Haim Levy. Risk Aversion and Skewness Preference. *Journal of Banking & Finance*, 32(7):1178–1187, July 2008.

Valerio Potì. The Coskewness Puzzle in the Cross-Section of Industry Portfolio Returns. Working papter, Dublin City University Business School, March 2006.

Valerio Potì and DengLi Wang. The Coskewness Puzzle. *Journal of Banking & Finance*, 34(8):1827–1838, August 2010.

Joshua V. Rosenberg and Robert F. Engle. Empirical Pricing Kernels. *Journal of Financial Economics*, 64(3):341–372, June 2002.

Mark E. Rubinstein. The Fundamental Theorem of Parameter-Preference Security Valuation. *Journal of Financial and Quantitative Analysis*, 8(1):61–69, January 1973.

William F. Sharpe. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, 19(3):425–442, September 1964.

Daniel T. Slesnick. Are Our Data Relevant to the Theory? The Case of Aggregate Consumption. *Journal of Business & Economic Statistics*, 16(1):52–61, January 1998.

Daniel R. Smith. Conditional Coskewness and Asset Pricing. *Journal of Empirical Finance*, 14(1):91–119, January 2007.

Pascal St-Amour. Benchmarks in Aggregate Household Portfolios. Cahiers de Recherches Economiques du Département d'Econométrie et d'Economie politique (DEEP) 07.07, Université de Lausanne, Faculté des HEC, DEEP, January 2007.

Dennis Stattman. Book Values and Stock Returns. *Chicago MBA: A Journal of Selected Papers*, 4:25–45, 1980.

Halbert White. A Heteroskedasticity-Consistent Covarince Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4): 817–838, May 1980.

David W. Wilcox. The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work. *American Economic Review*, 82(4):922–941, September 1992.

Alexandre Ziegler. Why Does Implied Risk Aversion Smile? *Review of Financial Studies*, 20(3):859–904, May 2007.

1.8 Proof of consistency of the pricing kernel estimation via factor model (OLS)

This appendix establishes the consistency of the estimation of the pricing kernel as estimated by the OLS factor model from section 1.3.2. For consistency, a more specific setup is required:

$$R_t^{e,k} - \mathbb{E}\left(R_t^{e,k}\right) = (\boldsymbol{f_t} - \mathbb{E}(\boldsymbol{f_t}))'\boldsymbol{\beta_k} + \epsilon_t^k \tag{1.12}$$

$$\mathbb{E}\left(R_t^{e,k}\right) = \boldsymbol{\beta_k'\lambda} + \eta_k^*, \tag{1.13}$$

where $R_t^{e,k}$ are the excess returns of asset k in period t, $\boldsymbol{f_t}$ is a stochastic vector of factors in period t, and the vector $\boldsymbol{\beta_k}$ is the factor exposures of asset k. The risk premia associated with the factors is the vector $\boldsymbol{\lambda}$. The vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\beta_k}$ for $k = 1, \ldots, K$ are fixed but with unknown parameters to estimate. ϵ_t^k and η_k^* are noise terms. The OLS estimator of $\boldsymbol{\beta_k}$ is $\boldsymbol{\hat{\beta}_k}$ and is estimated using:

$$R_t^{e,k} - \frac{1}{T}\sum_{t=1}^{T} R_t^{e,k} = \left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)' \boldsymbol{\beta_k} + \epsilon_t^k.$$

If the expected value is replaced with the sample average,[7] the second equation can be stated as follows:

$$\frac{1}{T}\sum_{t=1}^{T} R_t^{e,k} = \boldsymbol{\hat{\beta}_k'\lambda} + \eta_k.$$

[7]This replacement is unproblematic, as it is possible to include the estimation error only in the error term—i.e., $\eta_k = \eta_k^* - \mathbb{E}\left(R_t^{e,k}\right) + \frac{1}{T}\sum_{t=1}^{T} R_t^{e,k}$.

Using OLS, $\boldsymbol{\lambda}$ can be estimated from this equation by

$$\hat{\boldsymbol{\lambda}} = \left(\frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{R}_{\boldsymbol{t}}^{\boldsymbol{e}} \right)$$

with $\boldsymbol{R}_{\boldsymbol{t}}^{\boldsymbol{e}}$ as the vector of the excess returns of all assets in period t. A consistent estimator of the covariance matrix of $\boldsymbol{f_t}$ is

$$\widehat{\operatorname{Var}}(\boldsymbol{f_t}) = \frac{1}{T} \sum_{t=1}^{T} \left(\boldsymbol{f_t} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f_t} \right) \left(\boldsymbol{f_t} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f_t} \right)'.$$

From Section 1.3.2, it is known that the parameter of interest is $\boldsymbol{b} = -\operatorname{var}(\boldsymbol{f_t})^{-1} \boldsymbol{\lambda}$. An obvious candidate for an estimator of $\boldsymbol{b}$ is, therefore,

$$\hat{\boldsymbol{b}} = -\widehat{\operatorname{Var}}(\boldsymbol{f_t})^{-1} \hat{\boldsymbol{\lambda}}.$$

The next step is to show that $\hat{\boldsymbol{b}}$ is a consistent estimator. For this, the following assumptions are required:

- $\boldsymbol{f_t}$ is stochastic and $\left(\boldsymbol{f_t} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f_t} \right) \epsilon_t^k$ is a martingale difference sequence for all assets k
- $\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f_t} \xrightarrow{p} \mathbb{E}(\boldsymbol{f_t})$, where $|\mathbb{E}(\boldsymbol{f_t})| < \infty$
- $\mathbb{E}\left[\left(\epsilon_t^k\right)^2 \left(\boldsymbol{f_t} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f_t} \right) \left(\boldsymbol{f_t} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f_t} \right)' \right] = \boldsymbol{\Sigma_t}$, a positive definite matrix, with $\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\Sigma_t}$ converging to a positive definite matrix $\boldsymbol{\Sigma}$ and

$$\frac{1}{T} \sum_{t=1}^{T} \left(\epsilon_t^k\right)^2 \left(\boldsymbol{f_t} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f_t} \right) \left(\boldsymbol{f_t} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f_t} \right)' \xrightarrow{p} \boldsymbol{\Sigma}$$

- $\widehat{\operatorname{Var}}(\boldsymbol{f_t}) \xrightarrow{p} \boldsymbol{Q} = \operatorname{var}(\boldsymbol{f_t})$, where $\boldsymbol{Q}$ is nonsingular
- $\frac{1}{K} \sum_{t=1}^{T} \boldsymbol{\beta_k} \boldsymbol{\beta_k}'$ is a finite nonsingular matrix

- η_k and η_l, $l \neq k$ are independent random variables with $\mathbb{E}(\eta_k) = 0$ and $\text{var}(\eta_k) = \sigma_i^2 < \infty$.

- η_k is furthermore independent of all $\boldsymbol{f_t}$ and ϵ_t^k

The assumptions about $\boldsymbol{f_t}$ and ϵ_t^k are standard assumptions to ensure consistency for the parameters of a regression model with general heteroskedasticity in the error terms; this model dates back to Eicker (1967), White (1980), Hansen (1982) and Nicholls and Pagan (1983). Therefore, the OLS estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ for equation (1.12) converges to $\boldsymbol{\beta_k}$, that is,

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \xrightarrow{p} \boldsymbol{\beta_k}.$$

If $1/T \sum_{t=1}^{T} R_t^{e,k} = \boldsymbol{\beta}_{\boldsymbol{k}}'\boldsymbol{\lambda} + \eta_k$ is plugged into $\hat{\boldsymbol{\lambda}}$, the estimator $\boldsymbol{b}$ is

$$\begin{aligned}
\hat{\boldsymbol{b}} = -\widehat{\text{Var}}(\boldsymbol{f_t})^{-1}\hat{\boldsymbol{\lambda}} &= -\widehat{\text{Var}}(\boldsymbol{f_t})^{-1} \left(\frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}' \right)^{-1} \left(\frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \boldsymbol{\beta}_{\boldsymbol{k}}' \boldsymbol{\lambda} + \eta_k \right) \\
&= -\widehat{\text{Var}}(\boldsymbol{f_t})^{-1} \left(\frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}' \right)^{-1} \left(\frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \boldsymbol{\beta}_{\boldsymbol{k}}' \boldsymbol{\lambda} + \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \eta_k \right).
\end{aligned}$$

A first step to prove the consistency of $\hat{\boldsymbol{b}}$ is to show that

$$\frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \eta_k \xrightarrow{p} 0. \tag{1.14}$$

$\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ is a function of $\boldsymbol{f_t}$ and ϵ_t^k. η_k is independent of these two variables. Therefore, η_k and $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ are also independent and $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}\eta_k$ is a martingale difference sequence. The assumptions further imply bounded covariance matrices for η_k and $\boldsymbol{f_t}$. Therefore, $\text{cov}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}, \eta_k)$ is bounded. Then, by example 7.11 in Hamilton (1994), equation (1.14) holds. Further, $\widehat{\text{Var}}(\boldsymbol{f_t}) \xrightarrow{p} \text{var}(\boldsymbol{f_t})$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \xrightarrow{p} \boldsymbol{\beta_k}$. From the fact that there are only continuous functions in the estimator, it follows that

$$\hat{\boldsymbol{b}} \xrightarrow{p} -\operatorname{var}(\boldsymbol{f_t})^{-1} \left(\frac{1}{K} \sum_{k=1}^{K} \boldsymbol{\beta_k} \boldsymbol{\beta_k'} \right)^{-1} \left(\frac{1}{K} \sum_{k=1}^{K} \boldsymbol{\beta_k} \boldsymbol{\beta_k'} \boldsymbol{\lambda} + 0 \right)$$

$$\hat{\boldsymbol{b}} \xrightarrow{p} -\operatorname{var}(\boldsymbol{f_t})^{-1} \boldsymbol{\lambda} = \boldsymbol{b}.$$

In other words, $\hat{\boldsymbol{b}}$ is a consistent estimator for $\boldsymbol{b}$. However, the number of assets and the number of time steps have to converge to infinity for this estimator to be consistent.

Article 2

Firm Life Cycles under Financial Constraints and Additive Shocks

Joint work with Klaus Reiner Schenk-Hoppé

Abstract: This paper presents a simple model of the firm life cycle that captures several stylized economic and financial features which usually require considerably more demanding approaches. We study the optimal capital accumulation policy of a financially constrained firm whose revenue is subject to an additive shock. Earnings can be paid as dividends or reinvested with the goal to maximize shareholder value. In our model, the optimal policy of firms is to reinvest earnings (rather than paying dividends) when small, hold precautionary savings, and grow larger than is socially optimal. Smaller firms also have a higher bankruptcy risk and a more volatile market value than larger firms. We observe the leverage effect and excess returns of value stocks. In the presence of business cycles, investment and initial public offerings are pro-cyclical, the default probability is counter-cyclical, and monetary policy increases excess capital holdings but otherwise has a negligible impact.

2.1 Introduction

The theory of the firm life cycle, starting with the seminal contribution by Mueller (1972), continues to attract the interest of economists and finance researchers. At the heart of economic contributions to this theory is that a firm's investment opportunities and, therefore, its investment policy changes over time: young firms are innovative with high growth potential but lack capital; mature companies have few options for growth, face diseconomies of scale but are well capitalized.[1] From a financial perspective, the firm life cycle can be summarized as follows. Small firms are young, pay low (if any) dividends, grow quickly and have a high risk of bankruptcy while large companies are older, pay high dividends, barely grow and have a lower risk of default. Empirical support for the characterization of the life cycle through firms' dividend payments is given, e.g., by Fazzari et al. (1988), Fama and French (2001), Grullon et al. (2002) and DeAngelo et al. (2006). The growth/default perspective is supported by the findings of Hall (1987), Evans (1987a,b), Dunne et al. (1989) and Dhawan (2001).

Implications of financing constraints on the relation between the firm size and growth rates, default probabilities, and Tobin's q are discussed in Cooley and Quadrini (2001) who find that all of these measures are decreasing in the size of the firm. Other financial characteristics related to the size of the firm (and thus to the life cycle) are empirical observations on pro-cyclical investment behavior (Barro (1990)) and defaults (Chava and Jarrow (2004), Vassalou and Xing (2004)) and Chen (2010)).

This paper illustrates that these stylized empirical facts can be obtained in a very simple, neoclassical model where growth is purely driven by capital accumulation. To this end, we study the optimal dividend-investment policy of a firm whose earnings are subject to fluctuations in the output market (which acts as an additive shock to production). The firm does not have access to outside finance and growth has to be 'organic,' only accumulated capital can serve as a cushion against ex-

[1] Growth of a firm can take many different forms, for instance, development of new products or improvement in production efficiency through R&D, entering new markets, mergers and acquisitions to foster vertical or horizontal integration and many others.

ogenous shocks. Most effects are present under i.i.d. shocks, but we also consider the impact of the business cycle (in particular its depth and duration as well as central bank's interest rate policies) on the financially constrained firm's optimal behavior.

The role of financing constraints in the behavior of the firm and its implications for the dividend policy has been studied, e.g., by Cooley and Quadrini (2001), Albuquerque and Hopenhayn (2004), and Clementi and Hopenhayn (2006) whose results imply that credit restrictions can give rise to the firm life cycle.[2] Financing constraints force companies to save in order to ensure access to funding if and when needed. When companies do access their savings, however, is not obvious. Almeida et al. (2004) conclude that firms mostly save in times of high cash flows which enables them to realize investment opportunities in leaner times. Riddick and Whited (2009), in contrast, find that it is profit-maximizing to reduce savings in good times because these offer more profitable investment opportunities. In our model, the firm does not have access to any outside financing other than the initial investment by the owner; all growth has to be organic. Furthermore, financial market features as the value premium and the leverage effect can be explained by our model. This is in line with simulation results from Livdan et al. (2009).

Economics has produced a wealth of models explaining the firm life cycle of which only a few classical contributions are cited here. Mueller and Tilton (1969) discuss a technological-development cycle where many firms enter a new market and heavily invest in R&D which makes it more difficult for other newcomers to enter. Eventually technological progress slows and production techniques become standardized—the industry has matured and late entrants face large capital requirements. Therefore companies are growing fast (and face high risks of default) when they are young but their growth rate decreases over time. A similar dynamics occurs over the life cycle of a product, see, e.g., Gort and Klepper (1982) and Klepper (1996). A new product market draws in many entrants, reducing profitability and forcing exits by all but those who have the

[2]Financial constraints arise for a number of reasons, for instance, information asymmetries (Stiglitz and Weiss (1981), Greenwald et al. (1984), and Myers and Majluf (1984)) or agency conflicts (Jensen and Meckling (1976), Grossman and Hart (1982) and Jensen (1986)).

lowest R&D costs per produced unit—typically large companies. Other approaches stress the role of learning in firms' effort to determine their actual cost functions as, e.g., in the seminal contribution by Jovanovic (1982). In our model, there is only one firm with a fixed (and known) neoclassical production function whose earnings are subject to an additive shock and financial constraints.

In departure to the majority of literature as for example Cooley and Quadrini (2001) this paper uses additive shocks to the production function. This results in a non-concave value function and that companies may endogenously decide on the exit from the market (since a negative shock could wipe out the whole company, it can make sense to sell almost all assets, if a number of negative shocks are expected in future). Furthermore, additive shocks together with the financing constraint bring with them a surprising wealth of stylized facts in our simple model as shown in the literature discussion in the previous paragraphs.[3] An improvement to the existing literature is in that respect that the firm life-cycle and financial market effects as the value and the leverage effect can be explained in one model.

The remainder of this paper is organized as follows. Section 2.2 introduces the model. Section 2.3 numerically analyzes the firm's optimal dividend-investment policy for i.i.d. shocks and the resulting dynamic. Section 2.4 studies the impact of the business cycle on firms' behavior. Section 2.5 briefly looks into the effects of a central bank's interest rate policy. Section 2.6 concludes.

2.2 The model

We consider the optimal dividend-investment policy of a firm whose production is subject to exogenous shocks which entails random variations in earnings. The firm has no access to outside capital and growth has

[3]Our model is much simpler than the one of Cooley and Quadrini (2001), because no outside financing is allowed. This reduces the number of variables in the optimization problem as well as the total number of parameters in the model drastically. Due to the additive shocks on the other hand, the model gains more complexity.

to be organic. The firm can either retain profits to augment its capital stock, pay dividends to its owners, or combine both measures. The dividend payment stream is chosen such that its expected net present value is maximized. We consider the optimal dividend-investment policy of a firm whose production is subject to exogenous shocks which entails random variations in earnings. The firm has no access to outside capital and growth has to be organic. The firm can either retain profits to augment its capital stock, pay dividends to its owners, or combine both measures. The dividend payment stream is chosen such that its expected net present value is maximized. We assume that there are no agency conflicts between owner and management.

The firm's decision problem. Time is discrete with an infinite horizon, $t = 0, 1, \dots$. The shock $s \in \mathbf{S} := \{S_1, \dots, S_n\}$, $S_i \in \mathbb{R}$ for $i = 1, \dots, n$, follows a stationary time-homogeneous Markov process with transition probabilities $\pi_{s\tilde{s}}$, $s, \tilde{s} \in \mathbf{S}$. Given a state s, $E_s(\cdot)$ denotes the conditional expectation. We will study the firm's dividend policy first for an i.i.d. process (where $\pi_{s\tilde{s}}$ does not depend on s), and then for a proper Markov process (as a model of the business cycle).

The net production function $f(k, s)$ is assumed to be non-negative, continuous and bounded. We will further assume that the output $f(k, s)$ is increasing in the capital input investment) k and decreasing in the shock s. Our analysis will focus on production functions of the form

$$f(k, s) = \max\{g(k) - s, 0\}, \tag{2.1}$$

with a strictly concave function $g(k)$. If the output is zero (which happens when a shock of sufficiently large magnitude occurs), the firm is declared bankrupt because its output will remain zero in all future periods owing to a lack of access to outside finance.

The state of the shock is revealed after the capital is invested. Future payments are discounted with the discount factor $\beta \in (0, 1)$. The firm solves the following optimization problem for a given pair $(y_0, s_0) \in \mathbb{R}_+ \times \mathbf{S}$ of initial capital and initial state of the shock:

$$V(y_0, s_0) := \sup_{(d_t)_{t \geq 0}} \mathbb{E}_{s_0} \sum_{t=0}^{\infty} \beta^t d_t \tag{2.2}$$

subject to

$$y_{t+1} = f(k_t, s_{t+1}) \text{ and } 0 \leq d_t \leq y_t \text{ with } k_t := y_t - d_t. \tag{2.3}$$

Wasting capital is not optimal because the objective function is strictly increasing in each d_t. Therefore, the budget constraint in (2.3) is written as an equality. The above specification allows firms to pay out all initial capital as dividends in the first period without ever producing. The Bellman equation for the value of the optimization problem (2.2) is given by

$$V(y,s) = \sup_{0 \leq d \leq y} \left(d + \beta \sum_{\tilde{s} \in \mathbf{S}} \pi_{s\tilde{s}} V(f(y-d, \tilde{s}), \tilde{s}) \right). \tag{2.4}$$

Standard results (Stokey et al. (1989, Chapter 9)) ensure that there exists a unique solution $V(y,s)$ and a process $(d_t)_{t \geq 0}$ attaining the supremum in the optimization problem (2.2)–(2.3). Indeed, the supremum can be replaced by a maximum in (2.4). However, as the production function $k \mapsto f(k,s)$ is not necessarily concave, uniqueness of the optimal path cannot be guaranteed. We will choose the highest current dividend payment at which the maximum of the value function is attained. This selection rule leads to a unique dividend-investment policy.

Numerical approximation method. The numerical approximation of the value function uses the fact that the sequence

$$V_{n+1}(y,s) := \max_{0 \leq d \leq y} \left(d + \beta \mathbb{E}_s V_n(f(y-d, \tilde{s}), \tilde{s}) \right) \tag{2.5}$$

converges to the solution to (2.4), thanks to Blackwell's sufficient conditions for a contraction, see, e.g., Stokey et al. (1989, Theorem 9.6). The firm's optimal policy is determined numerically by solving the right-hand side of (2.5) for a given approximation of the value function. It suffices to approximate the value $V(y,s)$ on a set $[0, \bar{y}] \times \mathbf{S}$ with $\bar{y}$ sufficiently large because the net production function f and the set of shocks are bounded.

A firm without financing constraint. A useful benchmark is obtained by removing the restriction on access to outside financing. Suppose the firm can borrow and lend at an interest rate $r > 0$. The discount rate is $\beta = 1/(1+r)$. The optimization problem of the firm is unchanged but the budget constraint (2.3) is

$$y_{t+1} = f(y_t - d_t + b_t - (1+r)b_{t-1}, s_{t+1}), \text{ and } 0 \leq d_t \leq y_t + b_t - (1+r)b_{t-1}$$

with $b_{-1} = 0$. The principal amount b_{t-1} borrowed at time $t-1$ and the interest rb_{t-1} need to be repaid at time t. As debt can be rolled over, one has to assume that $\sup_t \mathbb{E} b_t < \infty$ to exclude Ponzi schemes.

The optimal investment $k^*(s)$, which depends only on the current state s, is determined by $\mathbb{E}_s[f'(k^*(s), \tilde{s})] = 1 + r$. Suppose that the production function is given by (2.1). Then, investing the capital k, the probability that there is no output in the current period is given by:

$$z_s(k) = \sum_{\{\tilde{s} \in \mathbf{S} \colon g(k) \leq \tilde{s}\}} \pi_{s\tilde{s}}.$$

The optimal investment $k^*(s)$ is given by the solution to:

$$k = (g')^{-1} \left(\frac{1+r}{1 - z_s(k)} \right). \tag{2.6}$$

If a solution $k^*(s)$ exists, the firm will operate forever and default will not happen because of access to outside financing in each period. Otherwise, the firm will not be established since its net present value would be negative. The additive shock is an extreme assumption, since the firm is hidden by the shock independent, if it is doing something or not. Especially the firm cannot decide to do nothing to be not exposed to the shock. The advantage of this assumption is that small firms are exposed to a relatively larger risk within a very simple structure.

Parameter values and production function. We consider a Cobb–Douglas type production function with negative shocks:

$$f(k, s) = \max\{\gamma k^\alpha - \delta k - s, 0\}, \tag{2.7}$$

where $\gamma > 0$, $0 < \delta < 1$ and $0 < \alpha < 1$. The values are set to

$$\alpha = 0.8,\ \beta = 0.95,\ \gamma = 2.0, \text{and } \delta = 0.1, \tag{2.8}$$

ensuring that small companies do not grow too fast (α is close to one to reduce marginal productivity for small capital stocks) and that firms with little capital are worth founding (γ is sufficiently large to avoid the optimality of paying all capital as dividends, without ever producing). Our focus is on the stylized features of the dynamics; no attempt is made to calibrate this simple model.

2.3 Optimal policy of the firm under i.i.d. shocks

The optimal dividend-investment policy of the firm is first studied for i.i.d. shocks. The production shock $s_t \in \{S_1, S_2\}$ with $S_1 = 0$ and $S_2 = 2$ is independent and identically distributed (i.i.d.) and assigns the same probability to each state, $\pi_1 = \pi_2 = 0.5$ (all transition probabilities $\pi_{s\tilde{s}} = 0.5$). The current state of the shock has no impact on the distribution of the next state and the value function is independent of s. The model parameters are defined in (2.7)–(2.8). In state S_1, production is standard Cobb–Douglas, with sustainable positive levels of capital stock. The state S_2, however, has a severe impact on the firm's capital and will deplete, after a long enough run, any amount of capital.

The value function is approximated numerically on a grid of 20,000 equidistant points in the interval $[0, \bar{y}]$, $\bar{y} = 20.0$. This set is forward invariant under the dynamics because $\max_{s \in \mathbf{S}} f(\bar{y}, s) = f(\bar{y}, 0) < \bar{y}$, i.e., no firm will accumulate more capital than $\bar{y}$. The numerical iteration (2.5) is performed until two subsequent functions are closer than 10^{-4} in the supremum norm $\|V\| = \sup_{0 \leq y \leq \bar{y}} |V(y)|$. From this approximation of the value function, we extract the optimal policy using the right-hand side of (2.5) subject to choosing the highest current dividend payment if the optimal decision is not unique. No numerical instabilities were encountered.

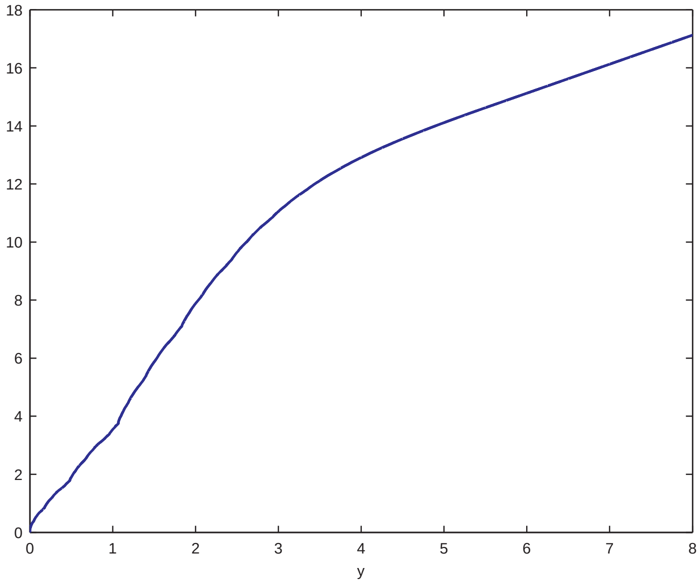
2.3.1 Dynamics

The firm's optimal investment, optimal dividend payment and the value function for given capital stocks are depicted in Figures 2.1(a)-2.1(c). The properties of each of these functions and their economic and financial implications are discussed in turn and compared with stylized empirical findings.

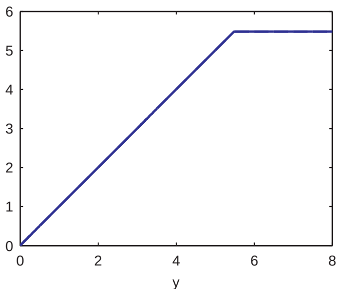
Value function. The value function, Figure 2.1(a), is not concave and exhibits kinks. This is a consequence of the non-smoothness of the production function for the shock $s = S_2$ because the risk of bankruptcy (i.e., the loss of the entire capital) does not depend continuously on the capital stock but rather jumps at levels of the capital stock which are exactly depleted when a run of n negative shocks prevails. Increasing capital at any of these critical levels by an arbitrarily small amount, drastically reduces the (n-run) bankruptcy risk. Although non-smoothness of the value function seems to have no effect on the optimal dividend policy, Section 2.3.3 shows that this is not true in general.

Dividend and investment policy. The optimal investment and dividend policy are depicted in Figures 2.1(b) and (c). The firm's policy is simple: below a certain capital level, all earnings are reinvested and no dividends are paid. If the output exceeds this threshold, then the capital stock is held constant and all 'excess earnings' are disbursed to the owners.

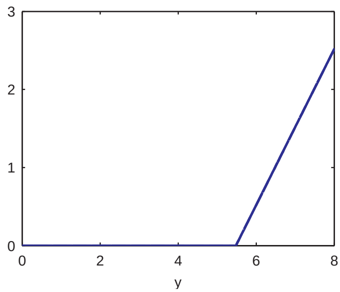
The threshold capital stock above which dividend payments are made is given by $k^* = 5.48$. This level is about 6.4% higher than the optimal (constant) investments of 5.15 which would be employed by a firm without financing constraints in every period. The higher capital stock reduces bankruptcy risk and allows the firm to rebuild its capital faster and resume dividend payments earlier, after the occurrence of the shock S_2. The firm's policy can be interpreted as precautionary savings which enable the firm to (temporarily) mitigate the effect of the shock. Cooley and Quadrini (2001) find the same investment policy in a model with default costs, costs of raising new capital and shocks that are proportional to the firm's output. In a continuous time setup with financial constraints and irreversible investment Holt (2003) finds a similar in-



(a) Value function.



(b) Optimal investment.



(c) Optimal dividend payment.

Figure 2.1: Value function $V(y_t)$, optimal investment k_t and optimal dividend payment d_t as a function of the initial capital stock y_t. Cobb–Douglas production function (2.7) with parameters (2.8) and uniformly distributed i.i.d. shock with values $S_1 = 0$ and $S_2 = 2$.

vestment policy: firms try to reach a certain optimal capital level, if the company is above that level it stops investing (in their setup the capital is irreversible, i.e. it cannot shrink).

In our model firms invest more than socially optimal and their optimal size is larger than if they had access to outside finance. Unlike in models with perfect capital markets such as Modigliani and Miller (1958) (where dividend payments can be offset by refinancing), the dividend policy matters and precautionary savings are optimal. The optimal policy of the firm matches empirical observations on retained earnings. Fazzari et al. (1988) find that firms with a value below 10 million dollars (small firms) have a retention ratio of 79%, whereas firms with a value over one billion have a ratio of 52%, i.e., smaller firms rely more on internal funding of investments. Guiso (1998), however, finds that size can be a poor proxy for measuring credit constraints.

Precautionary savings are, in practice, often related to holding more liquid assets, Opler et al. (1999). Our model makes no distinction between liquid and illiquid assets but the firm holds more assets than if it were unconstrained. Further evidence of precautionary savings is presented in Almeida et al. (2004) who find that companies save a larger proportion of their cash flow in good times (when the cash flow is high) in order to realize investment opportunities in times with low cash flows. This behavior closely resembles the precautionary savings observed in our model where the firm requires capital to survive negative shocks.

The firm's optimal dividend policy implies that larger companies pay more dividends and very small companies do not pay any. Fazzari et al. (1988) find that, in 1970, low dividend-paying firms were, on average, more than 12 times smaller than the high dividend-paying firms and that firms with low dividends, investments relative to capital are almost 50% higher than for high dividend payers. More recent findings by Fama and French (2001), and Grullon et al. (2002) are similar. According to DeAngelo et al. (2006), in 2003, only 18.9% of the companies paid any dividends.

Growth rates and default. The risk of default cannot be eliminated by the firm this parametrization, though the more capital that a company has, the longer it can survive. Therefore, the likelihood of default decreases with firm size, see Table 2.1. If a firm's capital is below 1.0671, it will default if the shock $S_2 = 2$ occurs. This default occurs independently of the investment decision because the shock will destroy the firm's entire capital within one period. A larger firm will survive longer.

The firm invests all of its capital if $y_0 \leq k^* = 5.48$. For $y_0 > k^*$, the firm pays its owners the amount $y_0 - k^*$ and invests k^*. As long as $y_t < k^*$, the firm aims to accumulate more capital. The firm grows if and only if the shock is $S_1 = 0$. When the shock S_2 occurs, the firm shrinks and continues with a lower capital stock in the subsequent period. A firm with capital stock of k^* can survive a run of 12 shocks of size S_2, but it would default after the thirteenth shock. (The probability of this event is $0.5^{13} \approx 0.0122\%$.) At the socially optimal capital stock ($k = 5.15$), the firm would default earlier.

	Current capital y					
	0.5	1.0	2.0	3.0	4.0	5.0
Prob. of default after 1 period	0.50	0.50	0.00	0.00	0.00	0.00
Prob. of default within 5 periods	0.75	0.59	0.22	0.03	0.00	0.00
Marginal productivity ($g'(y)$)	1.74	1.5	1.29	1.18	1.11	1.06
Exp. growth rate of capital (in %)	9.87	-5.00	14.11	17.21	16.57	14.69

Table 2.1: Probability of default over one and five periods, marginal productivity, and expected growth rate of the firm's capital for different current capital stocks y.

The relation between dividend payments and firms' growth rate is quite intricate in our simple model. Small (non-dividend paying) firms have a high marginal productivity but also a high risk of default, leading to a low expected growth rate (in the short run). Large firms have low marginal productivity and therefore grow slowly, if at all. Table 2.1 provides data on the expected value of $(k_1 - y_0)/y_0$, the expected growth rate of the firm size over $T = 1$ period. (In our model, positive growth only happens in the absence of the negative shock S_2.) Small as well as large companies experience falling growth rates with increasing size. The growth rate of smaller firms is more volatile because the the shock is independent of firm size; small firms are riskier than large firms.

Hall (1987), Evans (1987a,b) and Dhawan (2001) provide evidence, based on US manufacturing firm data, that small firms grow quicker, are more productive and riskier than larger firms. Models with financing constraints typically arrive at the same result, see, for example, Cooley and Quadrini (2001), Albuquerque and Hopenhayn (2004) and Clementi and Hopenhayn (2006). Beside that young and small firms grow quicker as Huynh and Petrunia (2010) showed, if the debt to asset ratio is low. I.e. deep pockets remove financial constraints and allow a faster growth. This is in line with our model, reducing financial constraints by providing more capital, increases the value of constraint firms massively. This can be seen especially by the steepness of the value function in figure 2.1(a) for small strongly constrained firms. The most extreme case is the unconstrained firm, which immediately jump to a much higher capital level.

Quantifying the effect of the financing constraint. The economic impact of the financing constraint on the firm's capital stock, compared with the social optimum, is twofold. Young companies are forced to grow organically and are forced to invest less than is socially optimal, whereas large companies hold too much capital as an insurance against bankruptcy risk. The loss to the shareholders incurred by the lack of outside finance is quantified in Table 2.2.

		Current capital y						
		0.0	0.5	1.0	2.0	3.0	4.0	5.0
NPV	Organic growth (1)	0.0	1.9	3.5	7.9	11.1	12.9	14.1
	Outside finance (2)	9.2	9.7	10.2	11.2	12.2	13.2	14.2
	(1) as % of (2)	0.0	19.6	34.3	70.5	91.0	97.3	99.3

Table 2.2: Net present value (NPV) of the firm as a function of the current capital stock: with and without financing constraint.

With access to outside finance, ownership of technology has value even if the firm does not hold capital; its net present value is 9.2. Increasing the initial capital, increases the value of the firm by the same amount. Without capital and outside financing, the firm has no value. As the initial capital paid by the shareholders increases, so does the value of the firm that needs to grow organically — but the effect is not linear. Small amounts of capital can have a large impact which is equal to the marginal productivity corrected for the risk of bankruptcy. The difference between the value of the firm in these two scenarios decreases with increasing equity. The financing constraint carries a substantial economic cost for smaller firms. This finding shows a large impact of financial constraints to the value and therewith also with the production of the firm. This is in line with the equilibrium model of Clemens and Heinemann (2010), who showed in a setup with labor market and intermediate goods calibrated to US data that tighter financial constraints lead to substantial losses in aggregate output and welfare.

2.3.2 Risk/return characteristics

The characteristic of risk and return profile of an investment into the firm matters, for instance, to the founder of a firm as well as to the investors when the firm goes public. We study an owner-entrepreneur who invests the initial capital y_0 and will be able to sell the firm for its net present value $V(y_T, s_T)$ in an IPO after T periods. (Here we assume that the investors are risk-neutral and the market is efficient.) The firm follows the optimal dividend policy described above.

Internal investment. The attractiveness of the initial investment into the firm can be measured by an average Tobin's q: The market value divided by the replacement value of the investment (Tobin (1969)). In our model, Tobin's q is given by $V(y)/y$, the net present value divided by the available capital y. This quantity describes the gross return to the firm's founder who invests y and immediately sells the firm for its net present value $V(y)$. Figure 2.2 depicts the relationship between Tobin's q and the initial capital y.

Tobin's q is high for small firms and decreases with firm size for larger capital stocks, though it is always larger than 1. Firms with a high Tobin's q reinvest all earnings, whereas those with a low Tobin's q pay dividends. The scope for expected future income, which can be realized with retained earnings, gives small firms a high value relative to its capital. Relative to the socially optimal size they are too small and the financing constraint bits particularly hard. These observations are

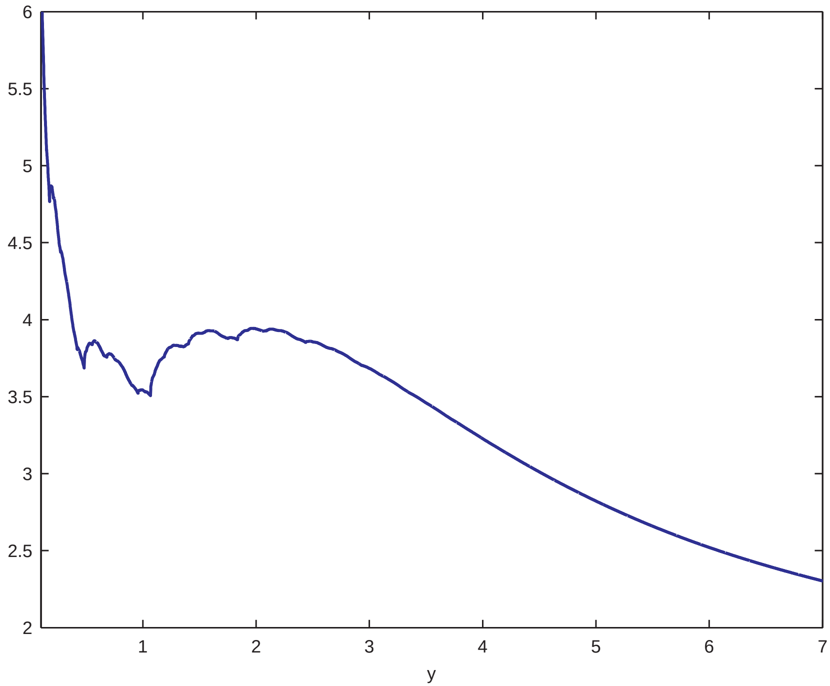


Figure 2.2: Tobin's q, $V(y)/y$, as a function of the initial capital y (i.e., the replacement value of the capital before dividend payment).

in lines with findings by Fazzari et al. (1988) and Erickson and Whited (2000) who show that constrained US firms have a higher Tobin's q and that these firms invest more. These properties also correspond to the simulation results presented in Cooley and Quadrini (2001).

For intermediate firm sizes, the non-concavity of the value function (owing to bankruptcy risk) implies a rather complex relationship which implies that, even in simple models, the relation between size and Tobin's q is not trivial. This puts into perspective the difficulties in finding strong empirical relations between Tobin's q and investment.

Outside investment. The gross return to a stock market investor who participates in the IPO is measured by the annualized ratio of the firm's net present value at time T and the ex-dividend net present value at time 0:

$$R_T(y_0, s^T) = \left(\frac{V(y_T(s^T))}{V(y_0) - d_0} \right)^{1/T}, \tag{2.9}$$

where $y_T(s^T)$ is the output in period T, which is determined by the sequence of shocks $s^T = (s_1, \ldots, s_T)$ and the firm's optimal dividend policy. The mean and variance of the return are given by

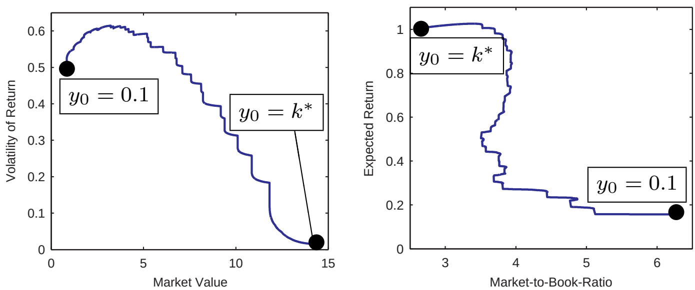
$$\mu_T(y_0) = \sum_{s^T \in \mathcal{S}^T} p(s^T) R_T(y_0, s^T)$$

and

$$\sigma_T(y_0)^2 = \sum_{s^T \in \mathcal{S}^T} p(s^T)(R_T(y_0, s^T) - \mu_T(y_0))^2,$$

where $p(s^T) = \pi_{s_1} \cdots \pi_{s_T}$ is the probability of observing the sequence of shocks s^T.

Empirical evidence of Whited and Wu (2006) and the simulation results of Livdan et al. (2009) suggest that, on average, more constrained firms have higher returns and higher volatility. In our model this holds only for very small firms. For example, if $T = 5$, the initial capital must be below $y_0 = 0.902$ which gives very unattractive expected return below $\mu_T = 0.44317$. Indeed, the return-volatility profile improves for investments up to $y_0 = 3.617$ where $\mu_T = 1.02608$ and $\sigma_T = 0.05425$.



(a) Market value and expected gross return.

(b) Market-to-book ratio and expected gross return.

Figure 2.3: Expected gross return $\mu_T(y_0)$ over $T = 5$ periods as a function of the firm's, market value $V(y_0)$ (panel (a)) and its market-to-book ratio $V(y_0)/y_0$ (panel (b)). Each point on the graphs corresponds to a particular initial capital stock y_0 with $y_0 = n \cdot 10^{-3}$, n an integer and $0.1 \leq y_0 \leq k^* = 5.48$.

For higher initial investments, the relation between the expected return and volatility follows the classical mean-variance diagram.

Leverage effect. As the company does not issue new capital, the market value of the firm is equal to its equity price and, therefore, some observations on equity returns can be made. Figure 2.3(a) presents the volatility of stock market returns, defined in (2.9), as a function of the value of the firm (i.e., its market capitalization). For large firms, volatility decreases with market capitalization. This effect is reversed for smaller firms with a value below 3.2240. Higher volatility as a result of falling equity prices, as observed for the large companies in Figure 2.3(a), is a stylized fact called the leverage effect. Black (1976) argues that a drop in the value of a company increases its leverage and, therefore, makes it riskier. Christie (1982) and Schwert (1989) show that volatility is an increasing function of leverage. The simulation results by Livdan et al. (2009) also show that financially constrained firms have a higher systematic risk.

Another possible explanation of the leverage effect is that a permanent increase in volatility increases, leads shareholders to demand a higher average return; therefore, today's price has to fall. This point is made, e.g., by Pindyck (1984), French et al. (1987), Turner et al. (1989), Campbell and Hentschel (1992), Wu (2001), Kim et al. (2004) and Mayfield (2004). Bekaert and Wu (2000) and Bae et al. (2007) quantify both effects in a model and find that the second effect is stronger. Since in the present model there is neither a risk premium nor leverage, our results show that these effects can be caused by financial constraints: A less valuable company becomes riskier because the likelihood of default increases if there is less capital to absorb shocks.

Value premium. The book value of a firm is the replacement value of the assets that the company owns. The market-to-book value ratio is an indicator of whether the company is a so-called 'growth' company or a 'value' company. A growth company (high market-to-book value) has few assets now, but the market expects the company to grow quickly and deliver substantial profit in the future. Value companies (low market-to-book value) already have many assets today, and their growth expectations are lower. Figure 2.3(b) illustrates the relation between the market-to-book ratio and expected returns in our model.

Firms with a low market-to-book value have a high capital stock but they also have high expected returns. The maximum expected return of 1.02608 is attained at a market-to-book ratio of 3.40557. Companies with little initial capital have a high market-to-book value but low returns and, by and large, the returns increase with a higher market-to-book ratio (see Figure 2.3(b)). This property is in line with the empirical findings by Stattman (1980), Rosenberg et al. (1985) and Chan et al. (1991). Fama and French (1992, 1998) and others found (using US and international data) that value stocks perform better than growth stocks, and that this effect cannot be explained by market risk factors. It is often argued that the premium for value stocks (i.e., stocks with a low book-to-market value) reflects other risk factors. In that view, the value premium is an indicator of the investment opportunities in the economy.[4]

[4]E.g., Fama and French (1996), Liew and Vassalou (2000), Campbell and Vuolteenaho (2004), Brennan et al. (2004), Hahn and Lee (2006) and Petkova (2006) or, for equilibrium models, Gomes et al. (2003), Zhang (2005) and Lettau and Wachter

Our model shows (as also observed by Livdan et al. (2009)) that these considerations are not needed if there are financial constraints. High market-to-book ratios may just be an indicator for financially constrained firms that also face a higher default risk, such that expected returns are lower.

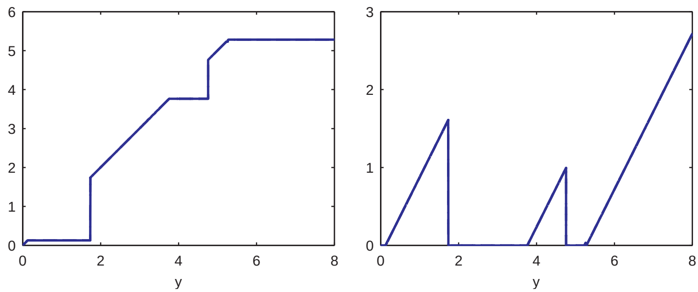
2.3.3 Multiple i.i.d. shocks

This section briefly studies an extended version of the previous model where the i.i.d. shocks can take on more than two values. It turns out that the non-concavity of the value function can indeed entail a more complex dynamics. For instance, the maximum size of the firm becomes a function of the initial investment. We assume that one additional (large) shock $S_3 = 6$ occurs with probability $\pi_3 = 0.09$ while the two other shocks $S_1 = 0$ and $S_2 = 2$ are assigned equal probabilities $\pi_1 = \pi_2 = (1 - \pi_3)/2$. Figure 2.4 shows the optimal investment and dividend as a function of the firm's capital.[5]

The optimal dividend-investment policy differs markedly from the one obtained in the previous case, displaying several 'plateaus' in Figure 2.4(a). In the present example there are three distinctive plateaus at different levels of capital stocks: low ($k_l^* = 0.128$), medium ($k_m^* = 3.766$) and high ($k_h^* = 5.283$). Each of these plateaus corresponds to a capital level above which the firm starts paying dividends. A firm with a capital stock below k_l^* pays out all earnings exceeding k_l^* and maintains this size until the shock S_2 or S_3 causes it to go bankrupt. The level k_h^* corresponds to the maximum size to which a firm can grow organically with an initial capital of $y_h^* = 1.739$ or more. For a capital level below y_h^*, the firm size converges to an optimal capital level of k_l^*. If the initial endowment is on the medium plateau between k_m^* and $\bar{k}_m^* = 4.761$, the firm will pay dividends of $y - k_m^*$ and shrinks to size k_m^*. If there is no negative shock (i.e., $S_1 = 0$ is realized), the company will grow in the next period from k_m^* directly to k_h^*. Therefore, the effect of the medium

(2007).

[5] Although the dynamics described below are prevalent in simulations with several shocks, parameters have to be chosen with a little care to obtain graphs as neat as those presented here.



(a) Optimal investment. (b) Optimal dividend payment.

Figure 2.4: Optimal investment k_t and optimal dividend payment d_t as a function of the firm's capital y_t for the case of three i.i.d. shocks.

level typically occurs only for one period. A firm of size k_h^* will retain earnings after one shock of size S_2, with the aim of reaching its previous size. If the large shock S_3 is realized, however, the capital stock falls below y_h^* and the firm will not grow to its previous size but rather shrink to size k_l^*.

2.4 Business cycles and optimal investment

The business cycle has a significant impact on firms' optimal dividend-investment policies. In this section we aim to study its effect within the framework of our model. The business cycle is implemented as exogenous market conditions with a certain degree of persistence, modeled by a Markov process with two shocks $S_1 = 0$ (boom) and $S_2 = 2$ (recession) and (symmetric) transition probabilities $\pi_{11} = \pi_{22} = p$. The probability of a change of the regime is given by $\pi_{12} = \pi_{21} = 1 - p$. The higher p, the higher is the persistence of a state and, thus, the average duration of regimes. The value function and the optimal policy of the firm will depend on the current state of the shock.

We are interested in qualitative differences in the firm's optimal policy between booms and recessions and, in particular, the effect of the duration of recessions (measured by p) and the depth (varying S_2) on the firm's optimal policy. The production function and parameters are given by (2.7)–(2.8) and are identical to the case studied in Section 2.3 (which is obtained by setting $p = 0.5$).

2.4.1 Optimal dividend-investment policy

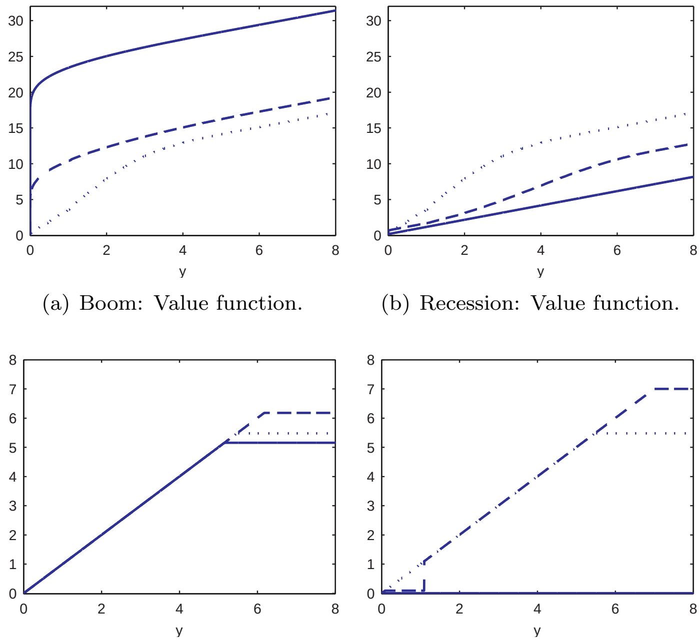
The value function $V(y, s)$ and the firm's optimal behavior is derived numerically from (2.5). The simulation results are presented in Figure 2.5.

The dividend policy in a boom is analogous to the i.i.d. case analyzed in Section 2.3. The firm reinvests all output up to a certain threshold, above which dividends are paid and the investment is kept constant. This threshold corresponds to the maximum size of the firm (i.e., the highest level of investment, which is reached after several boom periods). The maximum firm size in recessions, which is unobtainable but optimal for an outside investor, is larger than that in booms, see Figure 2.5(d).

Persistence p		0.50	0.75	0.90	0.95	0.975	0.99
Expected duration of regime		2	4	10	20	40	100
Maximum size (1)		5.48	6.07	6.18	5.69	5.15	5.15
Excess investment (in %)		6.36	17.77	19.85	10.38	0.00	0.00
Net present	$y = 1$	100	33.6	16.3	10.6	7.2	5.0
value of firm	$y = 2$	100	56.1	25.3	15.8	11.6	8.7
in recession	$y = 3$	100	70.3	35.5	21.4	15.6	12.1
(in % of (1))	$y = 4$	100	80.2	46.1	27.8	19.3	15.3
	$y = 5$	100	86.1	55.2	34.3	22.7	18.2

Table 2.3: Maximum size of the financially constrained firm in a boom and excess investment relative to an unconstrained firm (which holds capital 5.15 independent of the regime persistence). Net present value of the firm in the recession, $V(y, S_2)$, as a percentage of the value in the boom, $V(y, S_1)$, for different capital y.

Table 2.3 summarizes the impact of the persistence of regimes on the maximum firm size in booms and the difference in the firm value



(a) Boom: Value function.

(b) Recession: Value function.

(c) Boom: Optimal investment.

(d) Recession: Optimal investment.

Figure 2.5: The optimal policy of the firm in the two regimes for different degrees of persistence. Firm value $V(y_t, s_t)$ and investment k_t as a function of the capital y_t. Cases: $p = 0.5$ (dotted), $p = 0.9$ (dashed) and $p = 0.99$ (solid). Panels (a) and (c): current state $s_t = S_1 = 0$ (boom). Panels (b) and (d): current state $s_t = S_2 = 2$ (recession).

between booms and recessions (the firm's future prospects depend on the current regime and thus its valuation). The interpretation of the data on the maximum firm size is as follows. First, a firm without access to outside financing aims to accumulate excess capital as long as the regime persistence is not too high. The relationship between the duration of regimes and the maximum size is inverse U-shaped. The maximum amount of excess capital held is considerable (up to 20%). These precautionary savings prevent (or, at least, postpone) bankruptcy during a subsequent recession. This benefit decreases with the average duration of recessions which lowers the incentive to hold capital in excess of the socially optimal level.

The capital stock of the firms in our model is growing in good states (the companies invest) and decreasing in recessions. Barro (1990), using aggregated US investment data, shows that investment is pro-cyclical. Defaults in our model happen only in a recession which mimics the empirical fact that default probabilities are larger in periods of falling stock prices (Vassalou and Xing (2004) and Chava and Jarrow (2004)) which coincide with falling GDP (Chen (2010)).

The effect of regime duration on the variation of the firm value between booms and recessions is as follows, see Table 2.3. Smaller firms are harder hit by a recession. Whereas a large company will only lose about one-third of its value when entering a recession that lasts on average 10 periods, a small firm will see its value decline to 16.3%. This is caused by the low chances of survival in a persistent recession when the firm has little capital.

It is more attractive for an outside investor to invest during booms (since the same amount of capital y delivers a higher value $V(y, S_0) > V(y, S_1)$), implying that new companies will be founded mainly during booms. This feature matches the waves of IPOs documented by Ibbotson and Jaffe (1975) and Pástor and Veronesi (2005) who show that IPOs are more frequent in rising stock markets. A pro-cyclical pattern of firms' output, entry and exit is also found in the computational study by Delli Gatti et al. (2003). In their model firms face quadratic adjustment cost of capital and exogenously specified dividends. Companies default, if they are not able to pay the interest on their debts.

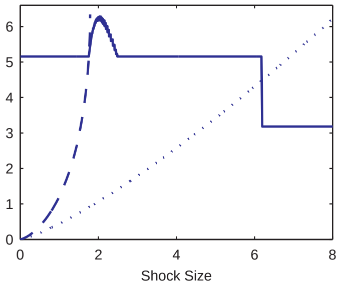
2.4.2 Effect of the depth of recession

We next study the effect of the depth of the recession on the firm's optimal dividend-investment policy. The depth is given by the size of the shock S_2 which will be varied in what follows (we set $p = 0.9$ and $S_1 = 0$).

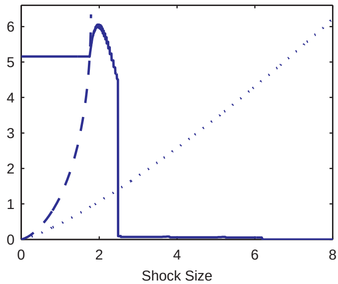
Bankruptcy eventually happens irrespective of the firm's policy if the shock is larger then the threshold $S_2 > 1.79$. For a smaller shock ($S_2 \leq 1.79$), there are sustainable levels of capital, i.e., $f(k, S_2) \geq k$. The size of the shock affects the firm's behavior during recessions as well as booms. Figure 2.6 shows the evolution of the firm's capital stock over time during a recession lasting five periods with the depth of the recession as a parameter. We assume that the firm enters the recession with the maximum capital level that it would attain in a boom.

The maximum size of the firm depends on the depth of the recession, as shown in Figure 2.6 (a). A company survives a recession of arbitrary length as long as its capital stock k is sustainable in a recession. The dashed lines in Figure 2.6 indicate the minimum capital level needed to avoid bankruptcy for a given shock size. Boundedness of the production function f implies that, for large shocks, bankruptcy is unavoidable. For the parametrization (2.7)–(2.8), the maximum size of the shock is 1.79. For larger shocks, it is certain that the firm will default sooner or later.

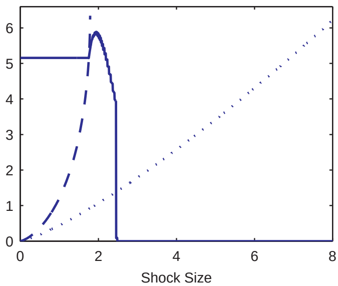
For smaller shock sizes, $S_2 \leq 1.76$, the firm accumulates capital up to the socially optimal level of 5.15 and, at this level, does not face any risk of bankruptcy. If the shock S_2 is larger, the firm holds more capital in booms. Holding this 'excess' capital reduces the risk of bankruptcy by postponing eventual bankruptcy during a recession because the firm has a capital buffer. This is illustrated in Figure 2.6 (b)–(d) by the 'hump' in the graph, which is located at shocks of size $1.76 \leq S_2 \leq 2.48$. The excess capital held by the firm first increases and then decreases with the depth of the recession. These precautionary savings are optimal because additional capital helps to postpone bankruptcy by many periods if the shock is just above the threshold 1.76. As the shock becomes larger, more excess capital is required to obtain the same benefit. At some shock size (here 2.12), the costs start to outweigh the benefit, leading to



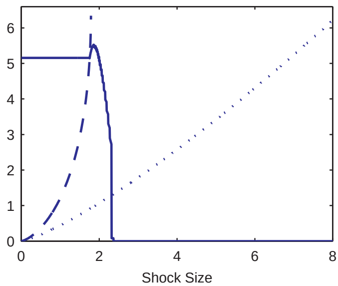
(a) Initial capital stock at $t = 0$.



(b) Capital stock at $t = 1$.



(c) Capital stock at $t = 2$.



(d) Capital stock at $t = 5$.

Figure 2.6: Capital stock as a function of the depth of the recession (size of shock S_2). Long-run capital stock in boom (a). Capital stock after being in recession for one, two and five periods: (b), (c) and (d). The firm enters the recession with the capital stock (a). Capital stocks above the dashed line are sustainable in a recession; capital stocks below the dotted line lead to bankruptcy if the next period is a recession.

lower precautionary savings.

If the shock size S_2 is larger than 2.48, the firm does not hold any excess capital in booms but rather chooses to accumulate capital up to the socially optimal level. When entering a recession, the firm pays out almost all of its remaining capital and keeps only a very small amount, which ensures survival only if a boom follows in the next period. Without instantaneous recovery, the firm will be bankrupt. If shocks are even more extreme (such that they would lead to the destruction of all of the firm's capital in one period of recession), the maximum size of the firm is 3.18. The inability to survive recessions induces firms to limit their size in the boom phase below the socially optimal one. This optimal behavior is evidenced as the downward step observed in Figure 2.6 (a) (and, although less visibly, in panel (b)).

2.5 Optimal policy of the central bank

We finally study the impact of the interest rate set by a central bank in response to prevailing economic conditions. The central bank is not able to anticipate the regime prevailing in the next period, but has to choose the interest rate $r(s_t)$ as a function of the currently observed state s_t. We denote by $r_1 := r(S_1)$ resp. $r_2 := r(S_2)$ the interest rate set in a boom resp. recession. (The policy will always lag the state of the economy by one period). We further assume that there is a given average interest rate r^* that the bank has to meet. In the symmetric case where, on average, booms and recessions last for the same number of periods, this condition can be written as $(1+r_1)(1+r_2) = (1+r^*)^2$. We will use the same specification of parameters as in Section 2.4 and set the persistence of regimes to $p = 0.9$. The interest rate $r^* = \beta^{-1} - 1 \approx 5.26\%$. We are interested in the central bank's optimal policy under these constraints.

The firm's decision problem is analogous to (2.2), with the discount factor β^t replaced by:

$$\beta^t(s^{t-1}) := \frac{1}{1+r(s_0)} \cdot \ldots \cdot \frac{1}{1+r(s_{t-1})}$$

for $t \geq 1$. Since the firm cannot hold debt, the interest rate policy has an impact on the firm's optimal decision only through the discount rate.

All results on the existence of the optimal dividend-investment policy and the value function remain valid. Table 2.4 summarizes the effect of the interest rate policy on the financially constrained firm's decisions.

	Standard policy			const.	Non-standard policy		
Boom interest rate r_1 (in %)	8.26	7.26	6.26	5.26	4.26	3.26	2.26
Recession interest rate r_2 (in %)	2.34	3.30	4.27	5.26	6.27	7.30	8.35
Constrained firm							
Maximum size in boom (1)	6.24	6.21	6.19	6.18	6.28	6.27	6.27
Minimum size in recession	1.54	1.40	1.24	1.10	1.07	1.07	1.07
Unconstrained firm							
Optimal capital in boom (4)	4.65	4.81	4.98	5.15	5.34	5.53	5.73
Difference (1) and (4) (in %)	34.08	29.11	24.37	19.85	17.64	13.37	9.34

Table 2.4: Impact of the central bank's interest rate policy on the size and market value of financially constrained firms in booms and recessions. Capital stocks of an unconstrained firm is provided as a benchmark.

The maximum size of the firm varies with the interest rate policy. The effect is quite unexpected because the maximum size of the firm increases with a non-constant policy irrespectively whether the 'boom interest rate' is increased or decreased. A higher interest rate during booms leads to stronger discounting of future boom dividend payments but lower discounting of payments during recessions. This makes precautionary savings more attractive because it enables the firm to survive longer in a recession. On the other hand, a higher interest rate in recessions (a rather unorthodox policy) has the effect that payments made during a recession is valued more by the owners of the firm than postponing it because of the higher discounting. This induces the firms to accumulate more capital in booms and pay high dividends in recessions, entailing an extremely high bankruptcy risk. The excess capital stock (last row in Table 2.4) held by the financially constrained firms in good times (booms) is increasing the lower the interest rate is set in recessions.

The minimum size (defined here as smallest the capital stock at which the firm chooses to continue operations in a recession rather than paying out almost all remaining capital) increases when the recession interest rate decreases. The effect of this standard interest rate policy, however, is driven by the high boom interest rate. In a recession the firm does not pay any dividends, their payment is only resumed in a boom which makes the capital holdings at the end of a recession more valuable. The firm therefore holds on to a higher capital stock in a recession.

Summarizing, the standard policy of low interest rates in recessions gives financially constrained firms an incentive to retain more of its earnings in good times (booms) and to stay longer in business in bad times (recession). In this sense, investments happen in booms out of precautionary motives. The persistence of regimes plays a vital role in the firm's decision, in the presence of i.i.d. shocks no adjustment to the business cycle would occur.

In the theoretical models of Bernanke and Gertler (1989), Kiyotaki and Moore (1997) and Carlstrom and Fuerst (1997), firms with limited access to the credit market (due to information asymmetries) cannot finance profitable investments because of macroeconomic shocks which reduce the value of the firms' collateral. The presence of this investment pattern is confirmed empirically by Gertler and Gilchrist (1994) and Bernanke et al. (1996). Monetary policy is found by Cooley and Quadrini (2006) to have a stronger (in terms of output and debts) impact of financially constrained small firms.

In our model, in contrast, financially constrained firms reduce their investment less than unconstrained firms at the outset of a recession because they accumulated precautionary savings in the previous boom. However, our results are about a different type of firm (those that do not have any access to credit) and, in addition, firms face persistent (rather than i.i.d.) shocks. We would argue that precautionary saving motives of financially constrained firm are more important under persistent recession regimes.

2.6 Conclusions

This paper studies the optimal behavior of a financially constrained firm in the presence of additive production shocks. The model is one of pure capital accumulation under i.i.d. as well as Markov (business cycle) shocks. Several stylized economic and financial characteristics of the firm life cycle can be illustrated within this simple model. The dynamic captures the higher default risk, productivity and volatility of small firms, the concentration of dividend payments on large firms, a falling Tobin's q in firm size, the leverage effect, the value premium, the pro-cyclically investment and firm entries, and counter-cyclical default probabilities. We also study the impact of a central bank's interest policy on firms' precautionary savings and their optimal size.

The approach offers several avenues for future research without losing much of the simplicity of the model. It would be interesting to weaken the (extreme) assumption of lack of access to any outside finance by allowing firms to raise at least some capital from, e.g., venture capitalists. The assumption on the risk-neutrality of the owner-entrepreneur can be replaced by other (neoclassical or behavioral) preferences. One could also study the impact of a proportional shock in the presence of fixed costs rather than imposing an additive output shock as in our model. Finally, competition of several firms in an output market (with some stochastic aggregate demand function) can also be studied in a generalized model.

Bibliography

Rui Albuquerque and Hugo A. Hopenhayn. Optimal Lending Contracts and Firm Dynamics. *Review of Economic Studies*, 71(2):285–315, April 2004.

Heitor Almeida, Murillo Campello, and Michael S. Weisbach. The Cash Flow Sensitivity of Cash. *Journal of Finance*, 59(4):1777–1804, August 2004.

Jinho Bae, Chang-Jin Kim, and Charles R. Nelson. Why Are Stock Returns and Volatility Negatively Correlated? *Journal of Empirical Finance*, 14(1):41–58, January 2007.

Robert J. Barro. The Stock Market and Investment. *Review of Financial Studies*, 3(1):115–131, March 1990.

Geert Bekaert and Guojun Wu. Asymmetric Volatility and Risk in Equity Markets. *Review of Financial Studies*, 13(1):1–42, Spring 2000.

Ben Bernanke, Mark Gertler, and Simon Gilchrist. The Financial Accelerator and the Flight to Quality. *Review of Economics and Statistics*, 78(1):1–15, February 1996.

Ben S. Bernanke and Mark Gertler. Agency Costs, Net Worth, and Business Fluctuations. *American Economic Review*, 79(1):14–31, March 1989.

Fischer Black. Studies of Stock Price Volatility Changes. Proceedings of the Meetings of the American Statistical Association, Business and Economics Statistics Division, 177–181, 1976.

Michael J. Brennan, Ashley W. Wang, and Yihong Xia. Estimation and Test of a Simple Model of Intertemporal Capital Asset Pricing. *Journal of Finance*, 59(4):1743–1775, August 2004.

John Y. Campbell and Ludger Hentschel. No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns. *Journal of Financial Economics*, 31(3):281–318, June 1992.

John Y. Campbell and Tuomo Vuolteenaho. Bad Beta, Good Beta. *American Economic Review*, 94(5):1249–1275, December 2004.

Charles T. Carlstrom and Timothy S. Fuerst. Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis. *American Economic Review*, 87(5):893–910, December 1997.

Louis K. C. Chan, Yasushi Hamao, and Josef Lakonishok. Fundamentals and Stock Returns in Japan. *Journal of Finance*, 46(5):1739–1764, December 1991.

Sudheer Chava and Robert A. Jarrow. Bankruptcy Prediction with Industry Effects. *Review of Finance*, 8(4):537–569, 2004.

Hui Chen. Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure. Journal of Finance, forthcoming, 2010.

Andrew A. Christie. The Stochastic Behavior of Common Stock Variances : Value, Leverage and Interest Rate Effects. *Journal of Financial Economics*, 10(4):407–432, December 1982.

Christiane Clemens and Maik Heinemann. On Entrepreneurial Risk-Taking and the Macroeconomic Effects of Financial Constraints. *Journal of Economic Dynamics and Control*, 34(9):1610 – 1626, 2010.

Gina Luca Clementi and Hugo A Hopenhayn. A Theory of Financing Constraints and Firm Dynamics. *Quarterly Journal of Economics*, 121 (1):229–265, February 2006.

Thomas F. Cooley and Vincenzo Quadrini. Financial Markets and Firm Dynamics. *American Economic Review*, 91(5):1286–1310, 2001.

Thomas F. Cooley and Vincenzo Quadrini. Monetary Policy and the Financial Decisions of Firms. *Economic Theory*, 27(1):243–270, January 2006.

Harry DeAngelo, Linda DeAngelo, and Rene M. Stulz. Dividend Policy and the Earned / Contributed Capital Mix: A Test of the Life-Cycle Theory. *Journal of Financial Economics*, 81(2):227–254, August 2006.

Domenico Delli Gatti, Mauro Gallegati, Gianfranco Giulioni, and Antonio Palestrini. Financial Fragility, Patterns of Firms' Entry and Exit and Aggregate Dynamics. *Journal of Economic Behavior & Organization*, 51(1):79–97, May 2003.

Rajeev Dhawan. Firm Size and Productivity Differential: Theory and Evidence from a Panel of US Firms. *Journal of Economic Behavior & Organization*, 44(3):269–293, March 2001.

Timothy Dunne, Mark J Roberts, and Larry Samuelson. The Growth and Failure of U.S. Manufacturing Plants. *Quarterly Journal of Economics*, 104(4):671–698, November 1989.

Timothy Erickson and Toni M. Whited. Measurement Error and the Relationship between Investment and "q". *Journal of Political Economy*, 108(5):1027–1057, October 2000.

David S. Evans. Tests of Alternative Theories of Firm Growth. *Journal of Political Economy*, 95(4):657–674, August 1987a.

David S. Evans. The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries. *Journal of Industrial Economics*, 35(4):567–581, June 1987b.

Eugene F. Fama and Kenneth R. French. The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47(2):427–465, June 1992.

Eugene F. Fama and Kenneth R. French. Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance*, 51(1):55–84, March 1996.

Eugene F. Fama and Kenneth R. French. Value versus Growth: The International Evidence. *Journal of Finance*, 53(6):1975–1999, December 1998.

Eugene F. Fama and Kenneth R. French. Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay? *Journal of Financial Economics*, 60(1):3–43, April 2001.

Steven M. Fazzari, R. Glenn Hubbard, and Bruce C. Petersen. Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity*, 1:141–195, 1988.

Kenneth R. French, G. William Schwert, and Robert F. Stambaugh. Expected Stock Returns and Volatility. *Journal of Financial Economics*, 19(1):3–29, 1987.

Mark Gertler and Simon Gilchrist. Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms. *Quarterly Journal of Economics*, 109(2):309–340, May 1994.

Joao Gomes, Leonid Kogan, and Lu Zhang. Equilibrium Cross Section of Returns. *Journal of Political Economy*, 111(4):693–732, August 2003.

Michael Gort and Steven Klepper. Time Paths in the Diffusion of Product Innovations. *Economic Journal*, 92(367):630–653, September 1982.

Bruce Greenwald, Joseph E. Stiglitz, and Andrew Weiss. Informational Imperfections in the Capital Market and Macroeconomic Fluctuations. *American Economic Review*, 74(2):194–199, May 1984.

Sanford J. Grossman and Oliver D. Hart. Corporate Financial Structure and Managerial Incentives. In *The Economics of Information and Uncertainty*, NBER Chapters, pages 107–140. National Bureau of Economic Research, Inc, August 1982.

Gustavo Grullon, Roni Michaely, and Bhaskaran Swaminathan. Are Dividend Changes a Sign of Firm Maturity? *Journal of Business*, 75 (3):387–424, July 2002.

Luigi Guiso. High-tech Firms and Credit Rationing. *Journal of Economic Behavior & Organization*, 35(1):39–59, March 1998.

Jaehoon Hahn and Hangyong Lee. Yield Spreads as Alternative Risk Factors for Size and Book-to-Market. *Journal of Financial & Quantitative Analysis*, 41(2):245–269, June 2006.

Bronwyn H. Hall. The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector. *Journal of Industrial Economics*, 35(4):583–606, June 1987.

Richard W. P. Holt. Investment and Dividends under Irreversibility and Financial Constraints. *Journal of Economic Dynamics and Control*, 27(3):467–502, January 2003.

Kim P. Huynh and Robert J. Petrunia. Age Effects, Leverage and Firm Growth. *Journal of Economic Dynamics and Control*, 34(5):1003–1013, May 2010.

Roger G. Ibbotson and Jeffrey F. Jaffe. ”Hot Issue” Markets. *Journal of Finance*, 30(4):1027–1042, September 1975.

Michael C. Jensen. Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *American Economic Review*, 76(2):323–329, May 1986.

Michael C. Jensen and William H. Meckling. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, 3(4):305–360, October 1976.

Boyan Jovanovic. Selection and the Evolution of Industry. *Econometrica*, 50(3):649–670, May 1982.

Chang-Jin Kim, James C. Morley, and Charles R. Nelson. Is There a Positive Relationship between Stock Market Volatility and the Equity Premium? *Journal of Money, Credit and Banking*, 36(3):339–360, June 2004.

Nobuhiro Kiyotaki and John Moore. Credit Cycles. *Journal of Political Economy*, 105(2):211–248, April 1997.

Steven Klepper. Entry, Exit, Growth, and Innovation over the Product Life Cycle. *American Economic Review*, 86(3):562–583, June 1996.

Martin Lettau and Jessica A. Wachter. Why Is Long-Horizon Equity Less Risky? A Duration-Based Explanation of the Value Premium. *Journal of Finance*, 62(1):55–92, February 2007.

Jimmy Liew and Maria Vassalou. Can Book-to-Market, Size and Momentum Be Risk Factors that Predict Economic Growth? *Journal of Financial Economics*, 57(2):221–245, August 2000.

Dmitry Livdan, Horacio Sapriza, and Lu Zhang. Financially Constrained Stock Returns. *Journal of Finance*, 64(4):1827–1862, August 2009.

Scott Mayfield. Estimating the Market Risk Premium. *Journal of Financial Economics*, 73(3):465–496, September 2004.

Franco Modigliani and Merton H. Miller. The Cost of Capital, Corporation Finance and the Theory of Investment. *American Economic Review*, 48(3):261–297, June 1958.

Dennis C. Mueller. A Life Cycle Theory of the Firm. *Journal of Industrial Economics*, 20(3):199–219, July 1972.

Dennis C. Mueller and John E. Tilton. Research and Development Costs as a Barrier to Entry. *Canadian Journal of Economics / Revue Canadienne d'Economique*, 2(4):570–579, November 1969.

Stewart C. Myers and Nicholas S. Majluf. Corporate Financing and Investment Decisions when Firms Have Information that Investors Do not Have. *Journal of Financial Economics*, 13(2):187–221, June 1984.

Tim Opler, Lee Pinkowitz, Rene Stulz, and Rohan Williamson. The Determinants and Implications of Corporate Cash Holdings. *Journal of Financial Economics*, 52(1):3–46, April 1999.

Lubos Pástor and Pietro Veronesi. Rational IPO Waves. *Journal of Finance*, 60(4):1713–1757, August 2005.

Ralitsa Petkova. Do the Fama-French Factors Proxy for Innovations in Predictive Variables? *Journal of Finance*, 61(2):581–612, April 2006.

Robert S. Pindyck. Risk, Inflation, and the Stock Market. *American Economic Review*, 74(3):335–351, June 1984.

Leigh A. Riddick and Toni M. Whited. The Corporate Propensity to Save. *Journal of Finance*, 64(4):1729–1766, August 2009.

Barr Rosenberg, Kenneth Reid, and Ronald Lanstein. Persuasive Evidence of Market Inefficiency. *Journal of Portfolio Management*, 11(3): 9–16, Spring 1985.

G. William Schwert. Why Does Stock Market Volatility Change Over Time? *Journal of Finance*, 44(5):1115–1153, December 1989.

Dennis Stattman. Book Values and Stock Returns. *Chicago MBA: A Journal of Selected Papers*, 4:25–45, 1980.

Joseph E. Stiglitz and Andrew Weiss. Credit Rationing in Markets with Imperfect Information. *American Economic Review*, 71(3):393–410, June 1981.

Nancy L. Stokey, Robert E. Lucas, and Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, 1989.

James Tobin. A General Equilibrium Approach to Monetary Theory. *Journal of Money, Credit and Banking*, 1(1):15–29, February 1969.

Christopher M. Turner, Richard Startz, and Charles R. Nelson. A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market. *Journal of Financial Economics*, 25(1):3–22, November 1989.

Maria Vassalou and Yuhang Xing. Default Risk in Equity Returns. *Journal of Finance*, 59(2):831–868, April 2004.

Toni M. Whited and Guojun Wu. Financial Constraints Risk. *Review of Financial Studies*, 19(2):531–559, Summer 2006.

Guojun Wu. The Determinants of Asymmetric Volatility. *Review of Financial Studies*, 14(3):837–859, Autumn 2001.

Lu Zhang. The Value Premium. *Journal of Finance*, 60(1):67–103, February 2005.

Article 3

Market Selection in an Evolutionary Market with Nonstationary Dividends

Abstract: Identifying investment strategies that will survive in the long run is a main endeavor in the field of evolutionary finance. The evolutionary perspective on the financial market considers rather long time horizons, making the creation and disappearance of firms a highly relevant factor in determining such strategies. However, this factor has not been examined in existing research. This paper seeks to fill the gap in the literature by simulating dividends and investment strategies on the basis of initial public offerings (IPOs) and defaults. This paper simulates the evolution of the wealth shares of various investment strategies in a setup wherein dividends are nonstationary. The results show that a modified version of the generalized Kelly rule dominates competing investment strategies in terms of final wealth. This finding agrees with the existing literature, which suggests that the generalized Kelly rule has good chances of surviving or even taking over the entire market in different setups. However, the creation and dissolution of a firm can only be observed once in the life of a company; therefore, using only a long time series of one company alone is not the most optimal method of estimating the probability that a firm will default. Instead, the dividend process must be understood by examining similar companies.

This completely alters the implementation of the generalized Kelly rule compared with the way it is applied in the existing evolutionary finance literature, even when the dividend processes of the companies involved are independent of each other.

3.1 Introduction

Financial analysis is based on the rationale that companies with similar characteristics exhibit a comparable firm value. One possible explanation for this may be that events that can only be observed once in the entire life of a company have a tremendous impact on the future of that company. Examples of such events are the first blockbuster product of a biotechnology company, the development of the iPhone by Apple Inc., or the default of a company. Moreover, such events can only be studied by examining similar companies. Further, this paper seeks to demonstrate that, in a market with several investment strategies, investors who incorporate cross-sectional information (i.e., information from other firms) in their investment decisions perform better than those who do not. In other words, the share of the total wealth accumulated using the strategies of the former increases more quickly than that accrued by those of the latter. Accordingly, strategies that perform poorly are marginalized in the long run. Briefly, this paper will show that the market selects investors who use cross-sectional information and that other strategies disappear in the long term.

The idea that the market selects investors who use all available information and who act rationally was initially proposed by Friedman (1953) and Fama (1965). According to these researchers, irrational investors earn lower returns and disappear in the long run. However, Long et al. (1990) used a partial equilibrium model to show that the effects of decisions made by irrational investors on stock prices cannot always be corrected by rational investors because the latter are risk averse. In addition, Blume and Easley (1992) proved that a rational investor who does not maximize a logarithmic utility function can be driven out of a complete market by some irrational investors, assuming that every investor has the same savings rate. For exogenous asset prices, Kelly (1956) developed a theory of maximizing expected returns on long-term (financial) investments. To do so, the investor has to bet his beliefs. Maximizing the growth rate as the Kelly rule does is equivalent to maximizing a logarithmic utility function. Therefore, a slightly irrational investor who almost maximizes a logarithmic utility function can push a rational investor who maximizes a nonlogarithmic utility function out of the market on the basis of the higher growth rate of his wealth as in Blume and Easley (1992).

Samuelson (1979) argued that individuals should maximize their utility (and therefore their happiness), regardless of whether they survive in the market. However, this paper focuses on identifying the strategies that survive a market selection process, rather than on making people happy. Because of the exogenously given savings rate, the asset allocations in Blume and Easley (1992) are neither Pareto optimal nor do they have a general equilibrium model. Sandroni (2000) and Blume and Easley (2006) investigated market selection in a general equilibrium setting with complete markets, and found that rational investors survive when all investors have the same discount rate, but the same does not apply in incomplete markets. These are addressed in detail by Evstigneev et al. (2006), and Evstigneev et al. (2008) showed that if all strategies and dividends possess Markov properties or that dividends are independent and identically distributed (i.i.d.), the generalized Kelly rule will drive all strategies that depend only on the actual state of the world out of the market, given that the initial wealth of the competing strategies is small enough (this property is called local evolutionary stability). This is further generalized by Amir et al. (2009b); they found that the generalized Kelly rule is asymptotically unique among all survival strategies that depend only on the history of states. This implies that the Kelly rule has almost surely a strictly positive wealth share that is independent of the strategy of the other investors. However, asset prices depend not only on the past states but also on the strategies of the other investors. Therefore, these results are for many relevant strategies as for example momentum strategies not applicable. However, simulation results from Tupak (2009) indicate that the generalized Kelly rule will dominate, given that the true parameters of the dividend process are known.[1]

[1]The generalized Kelly rule also applies to one-period assets with an arbitrary dividend process, see Evstigneev et al. (2002), Hens and Schenk-Hoppé (2005), and

What types of investment strategies survive if dividends are nonstationary and assets can be created and dissolved? Econometricians have long debated whether dividends contain a unit root or follow a stationary process, and the discussion is still not completely resolved.[2] Summarizing this debate, the test statistics of the unit root tests suggest that the hypothesis of a unit root in dividends is more difficult to maintain than the hypothesis of a unit root in stock prices. Further, Harris and Tzavalis (2004) have rejected the unit root hypothesis for dividends, and DeJong and Whiteman (1991) have also found it implausible. Therefore, this paper will concentrate on the implications of the second market feature, which states that assets can be created and destroyed. This feature automatically generates a nonstationary dividend process because many companies, including large ones that are very stable in the short term, did not exist, say, 200 years ago. To determine strategies that will survive in the long run, it is therefore important to consider the fact that companies can disappear and new companies will enter the market.

It is often assumed that dividends are driven by one and the same process over the entire life of a company. This is a very simplistic assumption: for instance, why should a small startup have the same risk and expected returns as a large concern? Mueller (1972) suggested a firm life-cycle: small firms are more profitable and face greater risks than large ones, but the large firms pay greater dividends. This life-cycle is driven by the idea that small firms tend to be more innovative but have difficulties in accessing the credit market and are therefore unable to pay dividends. Hall (1987) and Evans (1987a,b) found support for this theory in their work on US manufacturing firms, which led them to conclude that small firms grow more quickly and are riskier than larger firms. Similarly, Dhawan (2001) discovered that small US manufacturers are more productive and riskier than large ones. Fama and French (2001), Grullon et al. (2002), DeAngelo and DeAngelo (2006), and DeAngelo et al. (2006) provided empirical evidence for the life-cycle hypothesis of dividend policy, which holds that large firms pay more div-

Amir et al. (2005).

[2]For arguments surrounding the existence of a unit root in dividends, see Shiller (1981), Kleidon (1986), Campbell and Shiller (1987), Campbell and Shiller (1988), DeJong and Whiteman (1991), and Harris and Tzavalis (2004).

idends than small, growing companies. In addition to the possibility of default and the constructible and destructible nature of firms, this paper will consider the fact that small firms with small dividend payments may become large firms with high dividend payments.

The main aim of this paper is to find a surviving strategy for nonstationary dividends modeled on the creation and destruction of companies. Since the generalized Kelly rule is not feasible in this setup, the adopted strategy ensures that funds that are invested into a company are proportional to the expected net present value (NPV) of that companys dividends. This paper makes several observations. First, the NPV-strategy is able to dominate the markets in simulations; this indicates that the results from the infinitely lived assets seem to generalize to this setup. Second, many observations are required if the parameters of the process and the portfolio weights have to be estimated from past observations of the dividends. If past observations of dividends are lacking, a generalized Kelly rule with estimated parameters can be driven out of the market using simple strategies. This confirms the theoretical findings pertaining to infinitely lived assets obtained by Amir et al. (2009a) and the simulation findings for stationary dividends obtained by Tupak (2009). Third, if a wrong dividend process is assumed, the optimal Kelly rule can produce worse results than a naïve strategy that invests the same amount in each asset. This is shown in a case were the investor assumed the dividends to be i.i.d.,[3] while in reality, the dividends followed a nonstationary process.

Section 3.2 provides empirical evidence that the creation and disappearance of companies is an important factor in the dividend process and discusses further stylized facts. Section 3.3 presents a simple and minimalist dividend model that conforms to the literature discussed and the empirical evidence presented herein. Section 3.4 describes the market selection model in which the investment strategies detailed in Section 3.5 will compete. Section 3.6 simulates some models and Section 3.7 summarizes the papers findings.

[3] This assumption was used in the simulations of Hens et al. (2002), Hens and Schenk-Hoppé (2004), and Tupak (2009).

3.2 Empirical evidence on the birth, death, and dividends of companies

The present section motivates the assumptions for the dividend process described in the next section. This section mainly shows that, over the last 40 years, many new firms have been founded and are default, a fact that is often neglected in evolutionary finance. Furthermore, dividend payments are largely issued from a small number of companies, and the percentages of dividends paid by different sectors change over time. All of this demonstrates that dividends are highly nonstationary.

To illustrate these points, a sample of 25,272 active and inactive North American companies that are listed on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), or the NASDAQ Stock Market is used. The data have been obtained from the Center for Research in Security Prices (CRSP), and they are for the period 1973 through 2009. The starting date of 1973 was chosen because that is the earliest year from which AMEX and NASDAQ data are available. A company is considered active as long as it is listed on a certain stock exchange. Within this period, the number of active companies varies from 5,267 to 9,843 per year; the average is 7,138 companies per year. The large difference between the number of active companies per year and the number of active and inactive companies indicates that many firms were newly founded and that a large number of companies disappeared.

A company is active for an average time period of 10.51 years, and the median is lower: eight years. Figure 3.1 shows the number of companies that remain in the sample for any given number of years. Approximately a quarter of all firms were in the sample for less than four years, thus making it extremely difficult to determine the value of an asset on the basis of its past dividends. Therefore, cross-sectional information may be helpful in ascertaining the value and risk and therefore also the optimal amounts of investment in such assets.

The largest part of the variation in the number of active firms can be explained by mergers and acquisitions (see Table 3.1). Of the 18,837 delistings reported by the CRSP from 1973 to 2009, a total of 9,782 are

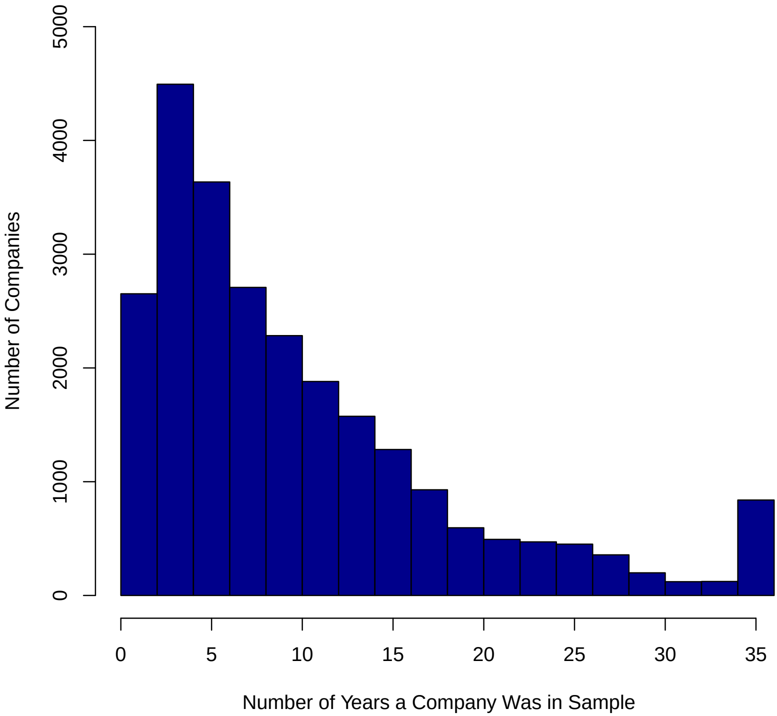


Figure 3.1: Number of years a company was in the sample (CRSP data from 1973 to 2009).

Year	Active	Mergers	Liquidation	Dropped
1973	5,954	103	5	354
1974	5,558	99	11	168
1975	5,398	83	10	80
1976	5,415	101	17	53
1977	5,390	155	18	61
1978	5,344	191	11	70
1979	5,267	209	16	51
1980	5,431	164	24	83
1981	5,794	155	15	88
1982	6,008	172	21	159
1983	6,606	173	11	164
1984	6,862	210	13	229
1985	6,982	248	16	325
1986	7,375	225	26	319
1987	7,642	185	5	225
1988	7,655	352	12	312
1989	7,390	273	13	318
1990	7,218	197	9	339
1991	7,251	119	12	367
1992	7,538	135	9	380
1993	8,108	177	3	170
1994	8,676	279	3	198
1995	9,055	358	9	233
1996	9,608	437	10	174
1997	9,843	511	7	255
1998	9,695	603	5	422
1999	9,374	613	11	394
2000	9,055	631	11	326
2001	8,337	463	7	473
2002	7,653	259	17	390
2003	7,228	257	11	297
2004	7,064	265	17	147
2005	7,043	268	7	162
2006	6,971	312	7	123
2007	7,000	392	8	184
2008	6,563	250	22	230
2009	6,237	158	36	267
Total		**9,782**	**465**	**8,590**

Table 3.1: Active companies (at the end of the year) and reasons for delisting (CRSP data from 1973 to 2009)

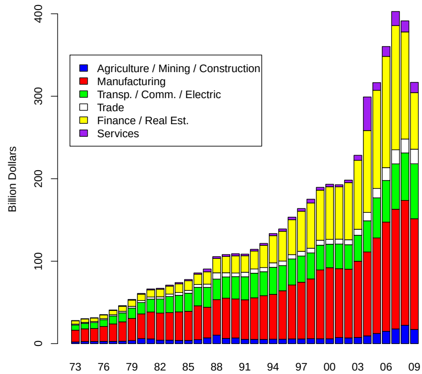
attributable to this source. The next most important reason for deletion is being dropped from the stock exchange. The number of dropped companies is much higher than the number of liquidated companies, and this indicates that the big stock exchanges delist companies with financial problems before the worst happens. Table 3.2 provides detailed reasons for dropping the companies: 1,282 companies were delisted because of insufficient capital, 930 because their price was too low, 647 because of insolvency, and 982 because they did not pay exchange fees. Therefore, the proportion of defaulting companies accounts for at least 12.5% of all companies, based on a time period of 36 years. The number of new companies is also significant: 19,318 such companies emerged during the period under study, working out to an average of 536.6 per year. These figures plainly demonstrate that long-run investment strategies should not neglect the fact that firms have finite lives.

Neither the numbers of delisted companies nor the reasons these companies were delisted are constant over time (see Table 3.1). Typically, everything happens in waves. For example, many new companies were founded between 1991 and 1997, and a merger wave occurred from 1996 to 2001. Between 1998 and 2004, the number of companies fell and the number of liquidations increased tremendously, and this phenomenon was repeated in 2008 and 2009. Due to this cyclical pattern, the number of companies also moves in waves. These findings correspond with the initial public offering (IPO) waves discovered by Ibbotson and Jaffe (1975), the procyclical behavior of IPOs noted by Pástor and Veronesi (2005), and the countercyclical behavior of default probabilities observed in Vassalou and Xing (2004), Chava and Jarrow (2004), and Chen (2010).

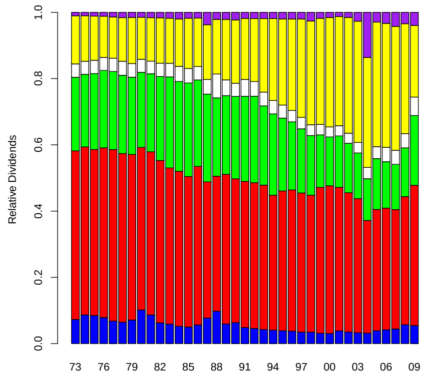
To aggregate dividends data from the CRSP, the number of outstanding shares on the day before the ex-distribution date must be multiplied by the dividend per share and then aggregated over one calendar year. For these calculations, only cash dividends were taken into account (i.e., subscription rights etc. were excluded from the estimations). Figure 3.2 aggregates the dividends by sector, and Figure 3.2(b) shows the relative weight of dividends being paid by different sectors. Until 2007, the dividends paid by the financial sector increased at a faster rate than those paid by the manufacturing industry and the transport and telecommunication sector. After the financial crisis in 2008, the dividends of the

Reason for Dropping	Firms
Issue stopped trading on current exchange–reason unavailable	965
Issue transferred from current exchange to Mutual Funds	18
Issue transferred from current exchange to Boston Exchange	33
Issue transferred from current exchange to Midwest Exchange	2
Issue transferred from current exchange to Pacific Stock Exchange	17
Issue transferred from current exchange to Philadelphia Stock Exchange	3
Issue transferred from current exchange to Toronto Stock Exchange	3
Issue began trading over the counter	375
Delisted by current exchange–insufficient number of market makers	464
Delisted by current exchange–insufficient number of shareholders	170
Delisted by current exchange–price fell below acceptable level	930
Delisted by current exchange–insufficient capital, surplus, and/or equity	1,282
Delisted by current exchange–insufficient (or noncompliance with rules of) float or assets	707
Delisted by current exchange–company request (no reason given)	512
Delisted by current exchange–company request (deregistration owing to going private)	81
Delisted by current exchange–bankruptcy (declared insolvent)	647
Delisted by current exchange–company request (offer rescinded and issue withdrawn by underwriter)	15
Delisted by current exchange–delinquent in filing and nonpayment of fees	982
Delisted by current exchange–failure to register under Section 12G of the Securities Exchange Act	112
Delisted by current exchange–failure to meet exception or equity requirements	167
Delisted by current exchange–denied temporary exception requirement	10
Delisted by current exchange–does not meet exchanges financial guidelines for continued listing	867
Delisted by current exchange–protection of investors and the public interest	137
Delisted by current exchange–corporate governance violation	13
Conversion of a closed-end investment company to an open-end investment company	47
Delisted by current exchange–required by the Securities Exchange Commission (SEC)	31

Table 3.2: Reasons that companies were dropped from their exchange (CRSP data from 1973 to 2009)



(a) Dividends by industrial sector



(b) Industrial sector dividends relative to total dividends

Figure 3.2: Dividends by Standard Industrial Classification (SIC) sector and SIC dividends relative to total dividends, that is, the dividends of one sector divided by total dividends (CRSP data from 1973 to 2009)

financial sector reduced drastically. This reveals persistent shifts in the relative weight of dividends being paid by different sectors. Such shifts are quite natural; the railroad and textile industries, for instance, were much more important one hundred years ago than they are today. Long-term shifts are incompatible with the assumption of i.i.d. dividend shares of the different sectors (or companies), and this assumption has often been made by parts of the evolutionary finance literature.

No dividends were paid in 57.2% of all company years, which are defined as the years during which a company is active. In 5.5% of all company years, dividends increased from zero to a positive amount and in 5.3% of the years, dividends fell to zero. This indicates that large variations in dividends are a very characteristic feature of dividend time series. Furthermore, the top 5% of dividend payers[4] distributed, on average, 78% of all dividends. This percentage varied between 71% and 85%, reaching its lowest in 1979 and peaking in 2001. These figures show that dividend payment is enormously concentrated among a few firms and that this concentration trended upward over time.

This section has noted several patterns relating to company numbers and dividend payment. To reflect these patterns, a long-run dividend model should exhibit the following features: a dynamic number of companies, wave-like changes in the number of companies, some large jumps in dividends, concentration of dividend payments among a small fraction of firms, and the capacity to accommodate shifts of dividends between different sectors.

3.3 Dividend model

This section will present a dividend model that is based on the stylized facts established in the previous section. Within this model, firms can be born and default, and a few large corporations issue a large percentage of dividend payments. In other words, small, young firms (startups) pay only a small dividend and are at high risk of defaulting, but have oppor-

[4] The top 5% of dividend payers constitute 5% of the companies paying the highest total dividends.

tunities to grow into concerns, which pay large dividends. To provide more detail, this model assumes an economy that consists of three types of companies: IPOs, startups, and concerns. IPOs are firms that have newly entered the market. In the entering period, investors pay a certain amount for an IPO and do not receive a dividend, and in the second period, the IPO automatically becomes a startup. Startups pay low dividends and may grow into concerns, which are mature firms that spend large sums on dividends but cannot grow further. These characteristics reflect the empirical evidence produced by Hall (1987), Evans (1987a,b), and Dhawan (2001), which demonstrates that younger firms have higher growth rates than older firms. Both types of companies can default and a company may change types in any period (a dead company being one such type). That is to say, a startup can default, remain a startup, or become a concern, and a concern can either default or remain a concern. A company that has defaulted is dead forever. The transition probabilities in Figure 3.3 are given as follows: p_{SD} is the probability that a startup will default during a period and p_{SS} is the probability that the startup will remain a startup. If the startup survives, then p_{SD} is the probability that it will default during the next period. The probability that a startup becomes a concern is then given by $p_{SC} = 1 - p_{SS} - p_{SD}$. A concern may remain a concern or default, and the probabilities for these events are p_{CC} and $p_{CD} = 1 - p_{CC}$.

Type changes are independent between companies and over time. If p_{SD} and p_{CD} are strictly positive, every company will default at one point in time, that is, if $t \to \infty$ the probability that a company is defaulted converges to 1. To guarantee that some companies will always exist, the number of IPOs in every period, n_{new}, exceeds zero. The number of startups, concerns, and IPOs are represented by n_t^S, n_t^C, and n_{new}, respectively. The long-run averages of the number of startups and concerns can be calculated as follows:

$$\mathbb{E}\left(n^S\right) = \frac{n_{new}}{1 - p_{SS}} \quad \text{and} \quad \mathbb{E}\left(n^C\right) = \frac{p_{SC} \cdot n_{new}}{(1 - p_{SS})(1 - p_{SC})}. \tag{3.1}$$

Every year, every startup pays a fixed strictly positive dividend D^S, and each concern pays a fixed dividend $D^C > D^S$. The fixed dividend D^S remains constant from the foundation of a startup to the point where

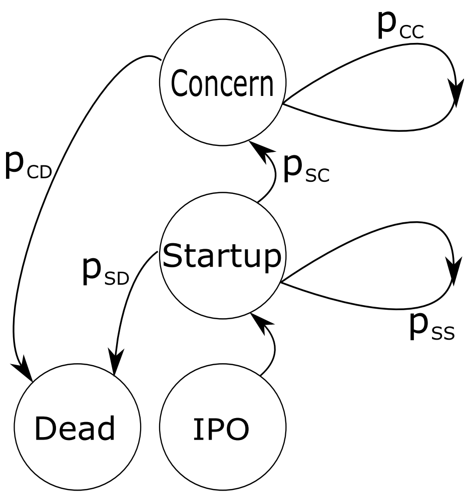


Figure 3.3: Development of a company over time. An IPO becomes a startup, a startup can become a concern or disappear after several periods, and a concern will exist for several periods and then disappear. The probability of each event is presented beside the arrows.

it either becomes a concern or dies. Should it become a concern, the dividend of the new concern experiences a huge upward jump; should it die, however, D^S permanently falls to zero. A concern pays D^C every year until it is dissolved. Therefore, the sum of all dividends paid by startups in a single period is the result of $n_t^S D^S$ and the total of all dividends paid by concerns during the same period is $n_t^C D^C$. Since IPOs do not pay dividends, it follows that the total dividend paid out in period t is the sum of the dividends paid by the startups and the concerns. The general structure of this dividend model is not completely new. Hurley and Johnson (1994) used a similar trinomial model to price individual stocks.

This model incorporates most of the features mentioned at the end of Section 3.2. The number of companies is dynamic and undergoes wave-like changes. Dividend payments can be parameterized so that they are largely paid by concerns and only fractionally disbursed by startups and are therefore concentrated among the concerns. To ensure that the simulation problem in Section 3.6 is tractable, the number of IPOs in every period is set to one. Simulations show that, with this assumption, the number of startups and concerns is changing drastically over time. Adding waves in the number of IPOs, as observed in the data of Section 3.2, would strengthen this effect further. Mergers are not included in the model because they do not matter, assuming that the dividends and portfolio weights of the new firm are the sum of the merging companies. Owing to the fact that new companies enter the market at all the times and existing companies disappear in the long run, changes in the dividends between several sectors can be explained by the model: in a certain time frame, mainly textile firms could enter the market, while in another period, only IT firms, and so on. That is, the sector of firms entering the market changes over time. After a certain period, the IPOs become concerns and pay considerable dividends, thereby leading to an increase in the importance of the sector. If no new firm of a certain sector enters the market, the sector disappears in the long run. As a whole, the model is very simple, but it includes many elements that are important in the long-run dividend process.

This paper aims to simulate the wealth shares of different investment strategies. For this, dividends must also be simulated. Table 3.3 shows

Startup		Concern		Miscellaneous	
D^S	1	D^C	40.1	n_{new}	1
p_{SD}	2.3%	p_{CD}	0.6%	λ_0	5%
p_{SS}	97.0%	p_{CC}	99.4%		
p_{SC}	0.7%				

Table 3.3: Simulation parameters calibrated on CRSP data

the parameters of the dividend process. These are calibrated with CRSP data for the years 1973 to 2009 in order to correlate the dividend process with the stylized facts underlying the dividend model. Table 3.2 does not confirm whether companies that were dropped from their exchanges were delisted owing to financial problems. Because of that uncertainty, the decision regarding which type a company belongs to is based on market capitalization: a company that has belonged to the top decile of all companies for at least two years is classified as a concern from that point until it defaults. Companies that neither qualified as concerns nor belonged to the lowest decile for at least two years in a row are classified as startups. A startup defaults if its value remains in the lowest decile for the rest of its life, whereas a concern defaults if its value remains below the largest 30% of all active companies for the rest of its life. The default threshold for concerns may seem high, but the market value of a concern defined as dead is approximately five times lower than the market value at which a startup becomes a concern. In other words, this threshold ensures that a concern must have suffered substantial losses before it defaults. The dividends of a startup, D^S, is normalized to one. The concern dividends, D^C, are determined in two steps. First, the quotient obtained by dividing every year the average of concern dividends through the average of startup dividends. This results in the dividends of the concerns in every year (D^S is normed to 1). Given that, the dividend of the concern is the average over all the years from which this quotient is derived. This dividend process is unrealistic for two reasons: first, it does not consider mergers, and second, it assigns equal dividend amounts to all company types. However, these simplifications allow us to observe the effects of the creation, growth, and default of firms on the wealth of investment strategies, which is the main purpose of this model.

3.4 Market selection model

As stated in the previous section, the dividend model is based on exogenously given dividends. In the next step of the dividend process, the companies generating these dividends are traded in a market and their shares may be purchased by several investors. This section will describe how the wealth of differing investment strategies and with that the asset returns evolve over time.

The state of nature in t is $\omega(t)$ and is described by the dividend payment of all companies at t. Therefore, the history of states equals $\omega^t = (\omega(0), \ldots, \omega(t))$. Given this, the percentage of wealth consumed by investment strategy i, at t in an economy with I investment strategies is defined as $\lambda^i_{0,t}(\omega^t)$, where $i \in \{1, \ldots, I\}$. This percentage is assumed to be constant over time and identical for all strategies because this paper focuses on comparing the performance of investment strategies, rather than analyzing the influence of the savings rate. Further, this assumption allows us to simplify $\lambda^i_{0,t}(\omega^t)$ to λ_0, which is important because it eliminates the possibility that an irrational strategy will survive by having a higher savings rate than rational strategies. In such a case, the rational strategy has a higher return, but the irrational strategy achieves a higher growth rate through a higher savings ratio and thereby marginalizing the rational strategy in the long run (see, e.g., Blume and Easley (1992)).

The percentage of wealth invested by shareholder i in company k at time t is represented by $\lambda^i_{k,t}(\omega^t)$. Nonexistent companies must have portfolio weights of zero; therefore, $\lambda^i_{k,t}(\omega^t) = 0$ for all companies that do not exist at t. Every strategy can invest in any existing company, but short selling is not allowed; that is, $0 \leq \lambda^i_{k,t}(\omega^t) \leq 1$. This budget constraint implies that $\sum_k \lambda^i_{k,t}(\omega^t) = 1$. Note that the sum over k can be interpreted as the sum over all past, actual, and future companies. However, it is not possible to invest in nonexistent companies, so this equates to adding up only the investments made in companies existing in period t. The wealth of investor i in t is w^i_t and the price of asset k in t is $q_{k,t}$. Therefore, the number of shares held by investor i in company

k at time t is

$$\theta^i_{k,t} = \begin{cases} \frac{\lambda^i_{k,t}(\omega^t)w^i_t}{q_{k,t}} & \text{if company } k \text{ exists in } t \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

If the number of stocks issued by a company is normalized to one and if all stocks need to be held by someone, the price of one stock in company k at t, or the market capitalization of that company, can be given as follows:

$$q_{k,t} = \sum_{i=1}^{I} \lambda^i_{k,t}\left(\omega^t\right) w^i_t =: \boldsymbol{\lambda_{k,t}}\left(\omega^t\right) \boldsymbol{w}_t \tag{3.3}$$

where $\boldsymbol{w}_t$ is a column vector including the wealth of all investors in period t and $\boldsymbol{\lambda_{k,t}}\left(\omega^t\right)$ is a row vector with the portfolio weights of all investors in asset k during period t. The asset prices can be written in matrix notation as follows:

$$\boldsymbol{q}_t = \boldsymbol{\Lambda}_t\left(\omega^t\right) \boldsymbol{w}_t, \tag{3.4}$$

where $\boldsymbol{q}_t$ is a price vector including all companies and $\boldsymbol{\Lambda}_t\left(\omega^t\right)$ is a matrix of all portfolio weights of period t. The number of rows represents the number of assets, and the number of columns represents the number of investors. Finally, the vector $\boldsymbol{w}_t$ includes the wealth of every investor in t. The wealth of an investor in $t+1$ is equal to the value of his portfolio plus the dividend payment:

$$w^i_{t+1} = \sum_k \left(D^k_{t+1}\left(\omega^{t+1}\right) + q_{k,t+1}\right) \theta^i_{k,t}. \tag{3.5}$$

The dividends of asset k at $t+1$ are represented by $D^k_{t+1}\left(\omega^{t+1}\right)$. Note that $q_{k,t+1}$ and $\theta^i_{k,t}$ depend on $\boldsymbol{w}_t$ and $\boldsymbol{\lambda_{k,t}}\left(\omega^t\right)$. The next step is to express the wealth dynamic in terms of exogenously given variables as the dividends and the strategy that depends only on the state of the world, ω_t. The preceding equation can be expressed in matrix notation as follows:

$$\begin{aligned}
\boldsymbol{w}_{t+1} &= \sum_k \left(D^k_{t+1}\left(\omega^{t+1}\right) + q_{k,t+1}\right)\boldsymbol{\theta}_{k,t}, && (3.6)\\
&= \boldsymbol{\Theta}_t \boldsymbol{D}_{t+1}\left(\omega^{t+1}\right) + \boldsymbol{\Theta}_t \boldsymbol{q}_{t+1}, && (3.7)\\
&= \boldsymbol{\Theta}_t \boldsymbol{D}_{t+1}\left(\omega^{t+1}\right) + \boldsymbol{\Theta}_t \boldsymbol{\Lambda}_{t+1}\left(\omega^{t+1}\right)\boldsymbol{w}_{t+1}, && (3.8)
\end{aligned}$$

where the last equation follows from equation (3.4). The number of shares held by all investors in all companies, $\boldsymbol{\theta}_{k,t}$, is a column vector with the length of the number of investors and $\boldsymbol{\Theta}_t$ combines the vectors $\boldsymbol{\theta}_{k,t}$ into a matrix in which the number of investors is designated by the number of rows and the number of assets by the number of columns. Furthermore, $\boldsymbol{D}_{t+1}\left(\omega^{t+1}\right)$ is the vector denoting the dividends of all assets in $t+1$, given the state ω^{t+1}. Writing $\boldsymbol{w}_{t+1}$ on one side of the equation results in

$$\left(\mathbf{I} - \boldsymbol{\Theta}_\mathbf{t}\boldsymbol{\Lambda}_\mathbf{t+1}\left(\omega^{\mathbf{t+1}}\right)\right)\boldsymbol{w}_{t+1} = \boldsymbol{\Theta}_t \boldsymbol{D}_{t+1}\left(\omega^{t+1}\right). \tag{3.9}$$

The evolution of wealth is therefore

$$\boldsymbol{w}_{t+1} = \left(\mathbf{I} - \boldsymbol{\Theta}_\mathbf{t}\boldsymbol{\Lambda}_\mathbf{t+1}\left(\omega^{\mathbf{t+1}}\right)\right)^{-1}\boldsymbol{\Theta}_t \boldsymbol{D}_{t+1}\left(\omega^{t+1}\right). \tag{3.10}$$

The next step is to check whether the wealth of all investors in $t+1$, $\boldsymbol{w}_{t+1}$, and the stock prices during the same period, $\boldsymbol{q}_{t+1}$, are always nonnegative. This is important because negative asset prices make no economic sense and negative wealth presents the problem of whether the investor will be able to pay back his or her debts. The system can be considered well-defined if the wealth of all investors and prices of all assets are nonnegative during all t. This requires three assumptions:

Assumption 1. *Consumption takes place and does not violate the rule that* $0 < \lambda^i_{0,t}\left(\omega^t\right) < 1$ *for all* i, t, *and* ω^t.

Assumption 2. *At least one completely diversified portfolio rule is in force: an i exists such that $\lambda^i_{k,t}(\omega^t) > 0$ for all existing k, t, and ω^t.*

Assumption 3. *If a company, k, is dead or not yet founded at t, then nobody invests in it (i.e., $\lambda^i_{k,t}(\omega^t) = 0$).*

Proposition 1. *Suppose that $w_0 > 0$ and assumptions 1 to 3 are satisfied. Then, the evolution of wealth* (3.10) *is well defined in all $t < \infty$.*

Mainly, the proposition holds because this setup is constructed so that both $\mathbf{\Theta}_t$ and $\mathbf{\Lambda}_t(\omega^t)$ contain many zeros; consequently, the step from t to $t+1$ is only influenced by companies existing in both periods. Considering this, the proof for Proposition 1 is analogous to that supplied by Evstigneev et al. (2006). Since the chief effect of Proposition 1 is to enable proper model simulation, the restriction to a finite number of time periods is not problematic.

The model may appear very similar to that of Evstigneev et al. (2006) or Amir et al. (2009a), but it is not possible to show that the generalized Kelly rule is locally evolutionary stable[5] or is a (unique) surviving strategy. A main prerequisite of their result is that consumption is a constant share of total wealth. In the present setup, shareholders pay a certain amount of money to establish a newly founded IPO, and this amount depends on the investment strategies of the investors and is therefore typically not constant over time. This fraction of investor wealth leaves the economy and is hence also a form of consumption, but because it is not constant over time, it is not possible to confirm the existence of principles such as local evolutionary stability or a unique surviving strategy. Results on these subjects are therefore provided by simulations. However, before this, the strategies to be considered for simulations must be defined.

[5] A strategy that drives every other strategy out of the market if the initial wealth of the other strategy is small enough is considered locally evolutionary stable.

3.5 Strategies

This paper has thus far delineated a dividend model and an evolutionary market selection model and will now proceed to discuss investor strategies, which must be known in order to simulate the entire market. Which strategies should compete in this model? A good starting point may be a generalized version of the Kelly rule:

$$\lambda^*_{k,t} = \frac{\lambda_0}{1-\lambda_0} \sum_{m=1}^{\infty} (1-\lambda_0)^m \, \mathbb{E}\left(d_{k,t+m}(\omega^{t+m}) \mid \omega^t\right),$$

where $d_{k,t}$ are the relative dividends of asset k at t (i.e., $d_{k,t} = \frac{D^k_t}{\sum_i D^i_t}$). The strategy λ^* has a probability of one of resulting in a positive wealth share when applied to both short (one-period) and infinitely long-lived assets see Amir et al. (2009a,b). However, neither this nor other results from literature dealing with local and evolutionary stability apply to the strategy λ^* in a market wherein new companies can be established or firms can default. Furthermore, the strategy is not directly applicable in the setup of this paper. This is because new companies will enter the market in future periods and will, therefore, pay positive relative dividends. These assets are not included in the calculation of λ^*, which must therefore not sum up to one. This issue can be circumvented by including only those companies that existed during t in the calculation of the relative dividends in the formula for $\lambda^*_{k,t}$. From a practical point of view, the lack of a closed-form solution for calculating λ^* is a more problematic issue. The strategy could be estimated via simulation, but doing so over an infinite time horizon would be time consuming and/or imprecise. Moreover, the evolutionary setup requires that this calculation be performed thousands of times, which was not practicable. Portfolio weights proportional to the NPV of the companies dividends provided a close substitute. The NPV of asset k with discount factor $1-\lambda_0$ is defined by

$$\mathrm{NPV}_k = \sum_{m=1}^{\infty} (1-\lambda_0)^m \, \mathbb{E}\left(D_{k,t+m}(\omega^{t+m}) \mid \omega^t\right).$$

The NPVs of the dividends of the different types of companies are as follows:

$$
\begin{aligned}
\mathrm{NPV}_{concern} &= \frac{(1-\lambda_0)\, p_{CC} D^C}{1-(1-\lambda_0)\, p_{CC}} \\
\mathrm{NPV}_{startup} &= (1-\lambda_0)\, \frac{p_{SS} D^S + p_{SC}\left(D^C + \mathrm{NPV}_{concern}\right)}{1-(1-\lambda_0)\, p_{SS}} \\
\mathrm{NPV}_{IPO} &= (1-\lambda_0)\left(\mathrm{NPV}_{startup} + D^S\right).
\end{aligned}
$$

If $NPV_{k,t}$ is defined as the NPV of asset k in period t, then the strategy based on the relative NPVs is as follows:

$$
\lambda^1_{k,t}\left(\omega^t\right) = \frac{\mathrm{NPV}_{k,t}}{\sum_j \mathrm{NPV}_{j,t}}.
$$

To find out whether the NPV strategy approximates the generalized Kelly rule, the portfolio weights of both strategies were calculated for several parameterizations and then compared. To determine the portfolio weights of the Kelly rule, the relative dividends of the companies were simulated 1,000 periods ahead, and the Kelly strategy was calculated using these dividends. This process was repeated 10,000 times. The average of these 10,000 realizations gives λ^*. This result not only shows that the formula for the NPV is similar to the formula for λ^*. In fact, the NPV strategy and the generalized Kelly rule are equivalent if total dividends in the economy are constant over time. The rest of the paper mainly uses one standard parameterization, which can be found in Table 3.3. The parameters of the standard parameterization are varied in Table 3.4, and the difference between the allocation of the Kelly rule, λ^*, and the NPV-strategy, λ_1 of these variations are presented in Panels A to D.[6] The total share of wealth invested in concerns, startups, and IPOs is calculated, and the percentage difference between the generalized Kelly rule and the NPV strategy is shown in Table 3.4. In most cases, the difference is well below 0.5%, which shows that the results produced by the two strategies are close to being identical. However, the differences between the two strategies widen massively when the default probability

[6] The parameters D^S, $p_{SS} = 1 - p_{SC} - p_{SD}$ and $p_{CC} = 1 - p_{CD}$ are not varied either because they are normed to one or given by the other parameters.

for concerns achieves 5% or more; overall, a generalized Kelly rule investor would invest almost 4% more in concerns than an NPV investor would under such circumstances. This indicates that differences exist between these two types of investors. Since the typical default probability for concerns is 1% or smaller, this difference has no impact on the simulations performed in the rest of this paper, wherein the NPV strategy is used as a proxy for the generalized Kelly rule because it can be calculated much faster than the latter can.

The previous strategy is called the theoretical NPV strategy since the parameters are assumed to be known. But in reality, the true parameters of the dividend model are unknown. Therefore, these parameters are estimated on the basis of past (simulated) data in order to compare this model with other models. D^C and D^S can be directly observed from the data, thereby making probability estimation quite simple. For example, p_{CD} can be estimated by dividing the number of defaults of concerns by the sum of the active concern years of all the concerns plus the number of defaults. This estimator is also a maximum likelihood estimator (MLE). With the estimated parameters, the competing strategy $\lambda^2_{k,t}\left(\omega^t\right)$ can be calculated in the same way as $\lambda^1_{k,t}\left(\omega^t\right)$, and it is called the empirical NPV strategy.

The next step is to find some interesting alternative strategies. Hens et al. (2002) and Hens and Schenk-Hoppé (2004) applied a simple strategy; they used average relative dividends as a proxy for λ^*. In the case of i.i.d. dividends, this is the Kelly rule. Therefore this strategy is

$$\lambda^3_{k,t}\left(\omega^t\right) = \frac{c_3}{\tau+1}\sum_{i=0}^{\tau}\frac{D^k_{t-i}\left(\omega^{t-i}\right)}{\sum_j D^j_{t-i}\left(\omega^{t-i}\right)} := \frac{c_3}{\tau+1}\sum_{i=0}^{\tau} d^k_{t-i}(\omega^{t-i}).$$

The number of periods over which averaging has been conducted is represented by τ and the factor c_3 is chosen such that $\sum_{k=0}^{K}\lambda^3_{k,t}\left(\omega^t\right) = 1$. This constant is needed because the environment of the existing companies differs in every time period and it is therefore not a given that the full budget of the agents is used after the averaging. This paper uses this strategy with the τ values of 100, 20, and 0. Considering a long history to estimate the relative dividends is effective in the case of i.i.d. relative dividends. In that case, the strategy converges to λ^*. This is

Panel A: Varying D^C

D^C		2	10	50	100
IPO:	$\lambda^*_{IPO} - \lambda^1_{IPO}$	0.021%	-0.001%	0.001%	-0.003%
Startup:	$\mathbb{E}\left(n^S\right) \cdot \left(\lambda^*_S - \lambda^1_S\right)$	0.156%	0.102%	0.165%	0.205%
Concern:	$\mathbb{E}\left(n^C\right) \cdot \left(\lambda^*_C - \lambda^1_C\right)$	-0.177%	-0.101%	-0.166%	-0.202%

Panel B: Varying p_{SC}

p_{SC}		10%	5%	2%	1%
IPO:	$\lambda^*_{IPO} - \lambda^1_{IPO}$	0.031%	-0.006%	-0.004%	0.000%
Startup:	$\mathbb{E}\left(n^S\right) \cdot \left(\lambda^*_S - \lambda^1_S\right)$	0.304%	0.258%	0.198%	0.214%
Concern:	$\mathbb{E}\left(n^C\right) \cdot \left(\lambda^*_C - \lambda^1_C\right)$	-0.336%	-0.252%	-0.194%	-0.215%

Panel C: Varying p_{SD}

p_{SD}		10%	5%	2%	1%
IPO:	$\lambda^*_{IPO} - \lambda^1_{IPO}$	-0.010%	-0.013%	-0.001%	-0.005%
Startup:	$\mathbb{E}\left(n^S\right) \cdot \left(\lambda^*_S - \lambda^1_S\right)$	0.100%	0.105%	0.120%	0.174%
Concern:	$\mathbb{E}\left(n^C\right) \cdot \left(\lambda^*_C - \lambda^1_C\right)$	-0.090%	-0.091%	-0.120%	-0.169%

Panel D: Varying p_{CD}

p_{CD}		10%	5%	2%	1%
IPO:	$\lambda^*_{IPO} - \lambda^1_{IPO}$	-0.144%	-0.120%	-0.019%	-0.009%
Startup:	$\mathbb{E}\left(n^S\right) \cdot \left(\lambda^*_S - \lambda^1_S\right)$	-3.585%	-3.637%	-0.430%	0.137%
Concern:	$\mathbb{E}\left(n^C\right) \cdot \left(\lambda^*_C - \lambda^1_C\right)$	3.728%	3.757%	0.449%	-0.127%

Table 3.4: Percentage difference between the Kelly rule, λ^*, and the NPV strategy, λ^1, in terms of total investment in IPOs, startups, and concerns. The parameters of the dividend process can be found in Table 3.3. In each panel, one parameter is varied. The strategies are calculated on the assumption that the number of IPOs, startups, and concerns is equal to their long-run averages of 1, $\mathbb{E}\left(n^S\right)$ and $\mathbb{E}\left(n^C\right)$. The generalized Kelly rule, λ^*, is obtained through 10,000 simulations over 1,000 time periods.

not the case in the selected dividend model, but this strategy is still an important benchmark. The case of $\tau = 0$ is special in that it relies only on the current relative dividends. Therefore, this strategy is the called current relative dividend strategy. Since it relies only on an extremely short history, it may be in a strong position in a setup where not much can be learnt from the past dividend history.

The last strategy diversifies naively; it invests the same amount into all existing assets. In other words,

$$\lambda_{k,t}^{5}\left(\omega^{t}\right)=\frac{1}{\text{Total number of active companies in } t}.$$

This strategy may appear somewhat unsophisticated, but DeMiguel et al. (2009) have showed that it performs astonishingly well on real data.

Obviously, many more strategies are possible. However, the generalized Kelly rule performs best in simulations within the i.i.d. and stationary setting, whereas mean-variance, adaptive, and even more sophisticated strategies have no chance of surviving (see Hens et al. (2002), Hens and Schenk-Hoppé (2004), and Tupak (2009)). Therefore, it makes sense to examine mainly those strategies on the basis of relative dividends, such as the generalized Kelly rule. To ensure that this inference holds, Section 3.6.3 compares the NPV strategy with a large number of fixed-mix strategies and confirms that the NPV strategy is not only a surviving strategy but, perhaps, also a locally evolutionary stable strategy.

3.6 Simulations

The main purpose of this section is to simulate the wealth dynamic of the competing investment strategies described above within the dividend process detailed herein. This will be done using equation (3.10). First, a simple example demonstrates the errors that can be produced by choosing an inadequate strategy by wrongly assuming a stationary dividend process. Second, simulations show that the NPV strategy is indeed able to take over a large part of the market, and finally, some robustness checks are performed on the NPV strategy.

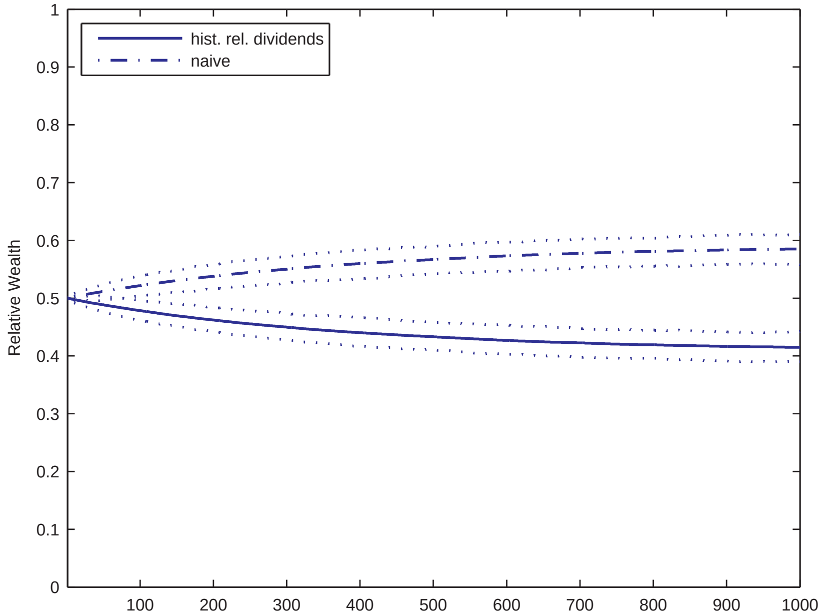


Figure 3.4: Relative wealth mean and 95% confidence intervals out of 1,000 simulations over 1,000 time periods. The market comprises two strategies: the historical relative dividend strategy averaged over the previous 20 time periods and the naive strategy. All strategies begin with equal wealth.

3.6.1 An illustrative example

With i.i.d. relative dividends, the Kelly rule λ^* is equal to the average past relative dividends, λ^3. This subsection shows that λ^3 is unable to dominate the market under the dividend model of this paper with nonstationary dividends and finitely lived assets, and it fails against the naive strategy of investing the same amount into each asset, λ^5. In contrast to our paper, most of the literature assumes i.i.d. dividends with infinitely lived assets for their simulations (see, e.g., Hens et al. (2002), Hens and Schenk-Hoppé (2004), or Tupak (2009)); this questions the relevance of these results. The dividend process is parameterized according to Table 3.3. To emphasize the results of this section, p_{SC} was set at 0.15 (as a consequence, $p_{SS} = 0.827$). This example establishes that a naive strategy that invests an equal share of wealth in every company can accumulate more wealth than a strategy based on the average relative dividends of the previous 20 periods (see Figure 3.4). The two strategies begin with equal wealth, and after 1,000 periods averaged out over 1,000 simulations, the naive $1/n$-strategy claims 58.5% of the total wealth, whereas the historical relative dividend strategy accounts for 41.5%.[7] [8] This result follows from the high probability that a startup will become a concern, which the relative dividend investor, who invests according to the past average relative dividends, neither knows nor takes into account because the event is observed once in the life of a firm. In contrast, the $1/n$-strategy increases its wealth share by investing more funds in startups than the relative dividend strategy does. Therefore, both strategies will survive in the long run. However, parameterizations calibrated on CRSP data show that the success of the naïve investor in the real world falls far short of the outcome achieved in this example. The simulated parameters in Table 3.3 put the real-world probability of a startup becoming a concern at a mere 0.7% (not 15%). This parameter results in an average wealth share of just 6.2% for the $1/n$-strategy after 1,000 periods. The simulations wherein $p_{SC} = 0.15$ are an example of how an optimal strategy may completely fail if a wrong dividend process is assumed. Therefore, it is extremely important for evolutionary simulations to assume a correctly specified dividend process.

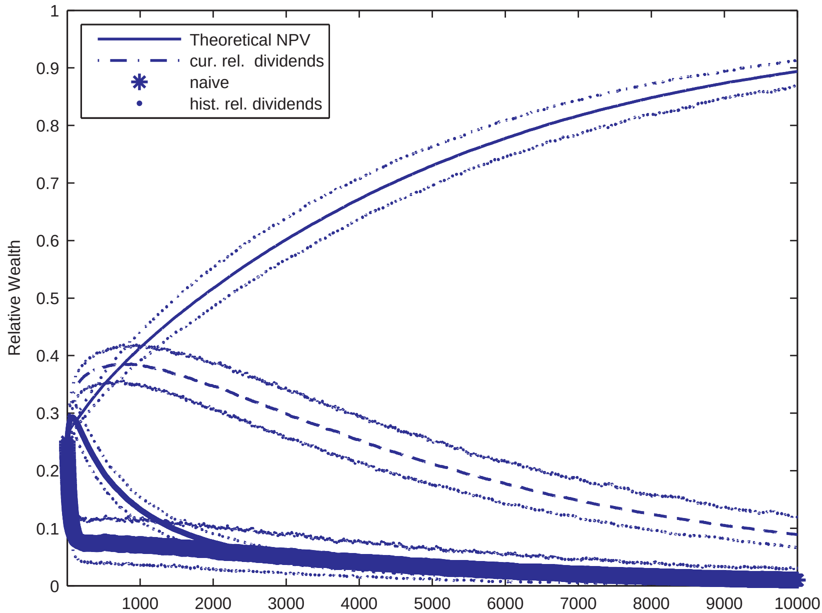


Figure 3.5: Relative wealth mean and 95% confidence intervals out of 1,000 simulations over 10,000 time periods. The market comprises four strategies: the theoretical NPV strategy, the current relative dividends strategy, the historical relative dividend strategy averaged over the previous 20 time periods, and the naive strategy. All strategies begin with equal wealth.

3.6.2 Is the NPV strategy able to take over the market?

The main purpose of this section is to demonstrate that an NPV strategy will finally take over almost the entire market. To achieve this objective, the parameters given in Table 3.3 were used to simulate changes in the wealth shares of the NPV, current relative dividend, historical relative dividend, and naive strategies. This involved estimating the mean and 95% confidence intervals of the relative wealth of the competing strategies on the basis of 1,000 simulations (see Figure 3.5). The results confirm that the theoretical NPV strategy outperforms all the other strategies to a striking extent. However, the strategies converge very slowly compared to those used by Hens et al. (2002) and Hens and Schenk-Hoppé (2004), who use i.i.d. dividend processes. In particular, the current relative dividend strategy loses wealth shares so slowly that after 1,000 periods, it still commands a larger market share than the NPV strategy does. Given that the model was calibrated so that one period equates to one year, this implies that evolutionary convergence can require an extremely long time horizon, especially if the competing strategies are not very dissimilar from the optimal one.

As mentioned earlier, a locally evolutionary stable strategy prevents invading strategies from earning higher returns when it has almost all the wealth in a market. Is the theoretical NPV strategy a locally evolutionary stable strategy? This question is easily answered via simulation. The process involves assuming that the theoretical NPV strategy begins with 97% of total wealth and that the three other strategies each start off with 1% of total wealth. On the basis of this assumption, 5,000 time periods are then simulated 1,000 times in order to determine whether the theoretical NPV strategy is able to retain its wealth share. The results of this procedure reveal that, after 5,000 periods, the theoretical NPV strategy owns an average of 97.7% of total wealth, with a standard

[7] Increasing the number of time periods to 5,000 (the default number of time periods in the rest of the paper) does not have a significant impact on the result.

[8] The simulations of the whole paper were also done with 100 simulations. The impact on average wealth, the main variable of interest, is minor. The only visible difference was that the confidence bands became smoother with 1,000 simulations. An even higher number of simulations is therefore unlikely to alter the results.

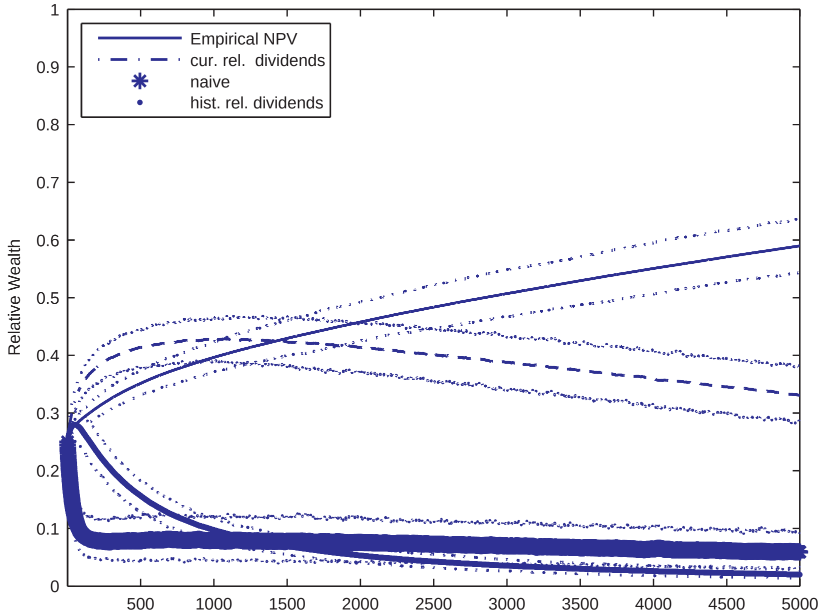
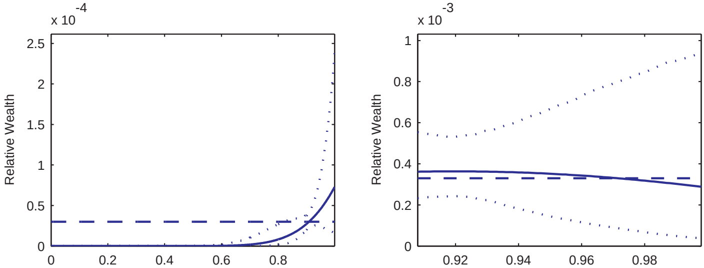


Figure 3.6: The relative wealth mean and 95% confidence intervals of 1,000 simulations of 5,000 time periods. The market comprises four strategies: the empirical NPV strategy, the current relative dividends strategy, the historical relative dividend strategy averaged over 100 periods, and the naïve strategy. All strategies start with equal wealth.

deviation of 0.9%. In contrast, the current relative dividend strategy accrues an average of 1.3% of total wealth with a standard deviation of 0.6%. The simulated distribution of the latters increase in wealth shows that it is not statistically significant at the 5% level. In other words, strategies with small total wealth shares are not able to wrest market share away from the theoretical NPV strategy. Therefore, the theoretical NPV strategy is evolutionary stable, at least against the chosen alternative strategies.

The wealth shares of the NPV strategy should also be determined using dividend parameters that are estimated from simulated dividend data, that is, the empirical NPV strategy. To this end, the parameters of the dividend process must be estimated over a sufficient number of time periods to ensure that the parameter values are precise enough. For example, if the parameters for calculating the empirical NPV strategy are determined on the basis of the previous 20 periods at every point of time, then the empirical strategy would be vanquished by the current relative dividend strategy. Starting with 97% of total wealth, the wealth share of the empirical NPV strategy would fall to an average of 57.0% of total wealth after 5,000 periods simulated 1,000 times each. Conversely, the wealth share of the current relative dividends strategy would expand from 1% to 28.5% of total wealth, and the historical relative dividends strategy and the naive strategy would gain 5.3% and 7.2% of additional wealth share, respectively. Figure 3.6 depicts the results of simulations using parameters obtained on the basis of the previous 100 periods. Both the empirical NPV and historical relative dividend strategies are now determined over the previous 100 periods so that the learning horizon remains consistent between them. In the same setup, the empirical NPV strategy acquires 98.0% of total wealth after 5,000 periods. All other strategies lose in wealth share aside from the relative dividend strategy, whose wealth share grows fractionally from 1% to 1.1%. In contrast to the results generated by parameters estimated from the previous 20 periods, these results show that the empirical NPV strategy easily conquers the other strategies when parameters calculated from the previous 100 periods are used. However, the empirical NPV strategy converges at a slower pace than it does in the simulations for the theoretical NPV strategy. This leads to the conclusion that, by definition, an imprecise estimation of the correct parameters affects the performance of the empirical NPV strategy, in some cases so much that the empirical NPV strategy has no chance to survive. This concurs with Tupak (2009), who finds that other strategies can perform better than λ^* if the latter must be learnt from the data. For infinitely lived assets and strategies that depend only on the state of the world, Theorem 2 of Amir et al. (2009a) implies that the optimal strategy learnt from the data must not survive against the optimal strategy that knows the true model parameters, that is, other strategies may triumph over the



(a) Fixed-mix strategies with an investment share of between 0% and 99.8% in concerns after 5,000 periods

(b) Fixed-mix strategies with an investment share of between 90.8% and 99.8% in concerns after 5,000 periods

Figure 3.7: The full line indicates the wealth share of fixed-mix strategies investing a constant share of wealth into concerns as averaged over 1,000 simulations. The dotted lines represent the 95% confidence interval of the strategies. The percentage of investment in concerns is shown on the x-axis. The percentage of investment in IPOs holds steady at 0.2% of the total investment and the remainder is invested in startups. The dashed line represents the initial wealth share.

estimated optimal strategy. The simulations by DeMiguel et al. (2009) showed that thousands of monthly observations are required before an optimal mean-variance strategy featuring asset returns that possess a multivariate normal distribution can overcome the naive $1/n$ strategy. This is mainly because the estimated average returns of the strategies contain a high level of error. Thus, the implementation of the theoretical optimal strategy may remain a challenge because of the errors in the estimation of the dividend process.

3.6.3 Robustness Checks

Obviously, the set of strategies in the market can greatly influence the outcome; therefore, the NPV strategy should be tested against as many other strategies as practicable in order to confirm the findings presented above. This was accomplished by further running the theoretical NPV strategy against a wide range of fixed-mix strategies. The fraction of funds invested in IPOs was assumed constant at 0.2% of total investment, which is the rounded average value of the NPV strategy if the number of companies is given by the long-run average defined in equation (3.1). Total investment in concerns was varied between 0% and 99.8% of total investment and was calculated using quantities that differed by 0.1% from each other. This investment was equally divided between all concerns and the rest of the investment was equally divided between all startups. This resulted in the creation of 998 different fixed-mix strategies to compete with the NPV strategies in the market. The theoretical NPV strategy started with 97% of total wealth with the rest divided equally among the fixed-mix strategies.[9] Figure 3.6.3 shows the average wealth share of the various fixed-mix strategies after 5,000 periods simulated 1,000 times each. The dashed line represents the average wealth share of these strategies in the first period. Fixed-mix strategies that invested less than 90.8% into concerns lose in average market share, while other fixed-mix strategies gain. Overall, the NPV strategy is able to increase its wealth share and ends up with 99.3% of total wealth. Naturally, the gains of the fixed-mix strategies that invested heavily in concerns are obtained at the expense of the fixed-mix strategies that invested limited amounts in concerns. To compare the successful fixed-mix strategies with the NPV strategies, all fixed-mix strategies that had gained in average wealth shares (i.e., those that invested 90.8% or more of their total wealth in concerns) were then matched against the NPV strategy. As before, all strategies involved were simulated 1,000 times per period over 5,000 periods. After 5,000 periods, the NPV strategy amasses a wealth share of 96.9%. The 95% confidence band rises from 93.9% to 98.6%, indicating that the initial wealth of the strategy does not differ statistically from the final wealth. In contrast, none of the fixed-mix strategies with a 95% confidence interval are able to gain a statistically significant proportion of wealth shares and their wins or losses more or less amount to zero (see Figure 3.7(b)). Overall, no fixed-

[9] This percentage was chosen to make the setup comparable with that delineated in Section 3.6.2. With an initial wealth of 90% of the theoretical NPV strategies, the results are comparable and the theoretical NPV strategy gains massively in wealth shares.

mix strategy is able to push the NPV strategy out of the market. On the other hand, the NPV strategy is also unable to push the fixed-mix strategies completely out of the market (although the wealth share of the latter is small). This evinces that the NPV strategy is a very close approximation to the real dominant strategy but is not itself the dominant strategy (given that such a strategy exists at all).

The results obtained in the previous section may differ according to variations in model parameters. Therefore, I ran additional tests in order to evaluate the main hypothesis that the NPV strategy is locally evolutionary stable compared with the current relative dividend, historical relative dividend, and naive investment strategies. This was done by estimating the aforementioned strategies with several different sets of parameterizations and examining the stability of the results thereby obtained. The NPV strategy was simulated with parameters calculated from the previous 100 periods and an initial wealth share of 97% and allocated the other strategies 1% each of wealth share. The results show increases in the wealth share of the empirical NPV strategy averaged out over 1,000 simulations over 5,000 time periods (see Table 3.5) and shows that this strategy can increase its relative weight under different parameterizations. This suggests that the NPV strategy can survive the evolutionary timeline with a large share of wealth and may therefore be at least close to a locally evolutionary stable strategy.

3.7 Conclusion

This paper demonstrates that large dividend jumps and the creation, growth, and default of companies are very important aspects of the dividend process. The idealized model presented herein shows that these factors have considerable influence on the performance of different investment strategies and signifies the inadequacy of considering only the time series of any given company in determining the percentage of wealth that should be invested into that company. Rather, these findings suggest that comparable companies should be studied in order to determine the optimal portfolio weight of companies. This is a new idea that complicates many aspects of evolutionary finance, including the estimation of

	Emp. NPV	Current rel. div.	Historical rel. div.	Naive investor
Initial relative wealth	0.970	0.010	0.010	0.010
Benchmark	0.980	0.011	0.000	0.008
	(0.002)	(0.001)	(0.000)	(0.002)
$\lambda_0 = 10\%$	0.982	0.012	0.000	0.007
	(0.002)	(0.001)	(0.000)	(0.002)
$d_C = 10$	0.975	0.013	0.002	0.010
	(0.002)	(0.001)	(0.000)	(0.002)
$d_C = 100$	0.984	0.009	0.000	0.007
	(0.002)	(0.001)	(0.000)	(0.002)
$n_{new} = 5$	0.978	0.012	0.002	0.008
	(0.001)	(0.000)	(0.000)	(0.001)
$p_{CC} = 98\%$	0.977	0.017	0.000	0.006
$p_{CD} = 2\%$	(0.008)	(0.005)	(0.000)	(0.006)
$p_{SC} = 2.0\%$	0.972	0.013	0.006	0.009
$p_{SS} = 95.7\%$	(0.002)	(0.001)	(0.001)	(0.001)
$p_{SD} = 5.0\%$	0.973	0.015	0.000	0.011
$p_{SS} = 92.3\%$	(0.004)	(0.002)	(0.000)	(0.003)

Table 3.5: The average relative wealth shares of the four main strategies after 1,000 simulations of 5,000 periods. The figures in parentheses express the standard deviations of the given percentages. The market comprises four investors: the empirical NPV investor (learning over 100 periods), the current relative dividends investor, the historical relative dividend investor averaging over 100 time periods, and the naive investor. Eight models were estimated: the benchmark model, which was calculated according to the parameterizations in Table 3.3, and seven other models wherein one parameter has been different compared to the benchmark model. The varied parameter and its new value can be found in the leftmost column.

an appropriate dividend model, and constitutes an important drawback: even very primitive strategies can outperform the most elaborate ones if a huge amount of data is required to estimate them accurately. Nevertheless, the NPV strategy, a close substitute of the generalized Kelly rule, dominates within this setup, although it requires long time periods to approximate a 100% wealth share. Alternatively, this result could also indicate that the strategy that is able to achieve the most precise calculations of the fundamental value of a firms dividends will be the one to survive or even dominate the market in the long run.

Simulation studies, such as this one, inherently face one major issue: it is never possible to test the whole range of possible parameters. Therefore, even with the extensive robustness checks carried out within the paper, there is no guarantee that the results found can be generalized for all cases. Furthermore, the Kelly strategy does not generalize to the chosen setup and must be approximated by the NPV strategy. Therefore, the present paper provides only a rough approximation of an evolutionary stable strategy.

This study generates three interesting directions for future research. First, theoretical results pertaining to nonstationary dividends and finitely lived firms would ascertain whether the results provided by simulations in this paper are generally applicable. Second, future work could perform simulations on the basis of alternative stochastic dividend processes in order to investigate the impact of such processes on the surviving strategy. Third, this paper has shown that the learning period may wield a crucial influence on strategy performance. Therefore, future work could attempt to determine how the dividend process should be learned optimally, such that the optimal strategy based on those results is able to take over the market.

Bibliography

Rabah Amir, Igor V. Evstigneev, Thorsten Hens, and Klaus Reiner Schenk-Hoppé. Market Selection and Survival of Investment Strategies. *Journal of Mathematical Economics*, 41(1-2):105–122, February 2005.

Rabah Amir, Igor V. Evstigneev, Thorsten Hens, and Le Xu. Evolutionary Finance and Dynamic Games. Working Paper 581, NCCR "Financial Valuation and Risk Management", September 2009a.

Rabah Amir, Igor V. Evstigneev, and Klaus Reiner Schenk-Hoppé. Asset Market Games of Survival. Working Paper 505, NCCR "Financial Valuation and Risk Management", September 2009b.

Larry Blume and David Easley. If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets. *Econometrica*, 74(4):929–966, November 2006.

Lawrence Blume and David Easley. Evolution and Market Behavior. *Journal of Economic Theory*, 58(1):9–40, July 1992.

John Y. Campbell and Robert J. Shiller. Cointegration and Tests of Present Value Models. *Journal of Political Economy*, 95(5):1062–88, October 1987.

John Y. Campbell and Robert J. Shiller. The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies*, 1(3):195–228, 1988.

Sudheer Chava and Robert A. Jarrow. Bankruptcy Prediction with Industry Effects. *Review of Finance*, 8(4):537–569, 2004.

Hui Chen. Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure. Journal of Finance, forthcoming, 2010.

Harry DeAngelo and Linda DeAngelo. The Irrelevance of the MM Dividend Irrelevance Theorem. *Journal of Financial Economics*, 79(2): 293–315, February 2006.

Harry DeAngelo, Linda DeAngelo, and Rene M. Stulz. Dividend Policy and the Earned / Contributed Capital Mix: A Test of the Life-Cycle Theory. *Journal of Financial Economics*, 81(2):227–254, August 2006.

David N. DeJong and Charles H. Whiteman. The Temporal Stability of Dividends and Stock Prices: Evidence from the Likelihood Function. *American Economic Review*, 81(3):600–617, June 1991.

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal Versus Naive Diversification: How Inefficient is the 1-N Portfolio Strategy? *Review of Financial Studies*, 22(5):1915–1953, May 2009.

Rajeev Dhawan. Firm Size and Productivity Differential: Theory and Evidence from a Panel of US Firms. *Journal of Economic Behavior & Organization*, 44(3):269–293, March 2001.

David S. Evans. Tests of Alternative Theories of Firm Growth. *Journal of Political Economy*, 95(4):657–674, August 1987a.

David S. Evans. The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries. *Journal of Industrial Economics*, 35(4):567–581, June 1987b.

Igor Evstigneev, Thorsten Hens, and Klaus Reiner Schenk-Hoppé. Market Selection of Financial Trading Strategies: Global Stability. *Mathematical Finance*, 12(4):329–339, October 2002.

Igor Evstigneev, Thorsten Hens, and Klaus Reiner Schenk-Hoppé. Evolutionary Stable Stock Markets. *Economic Theory*, 27(2):449–468, January 2006.

Igor V. Evstigneev, Thorsten Hens, and Klaus Reiner Schenk-Hoppé. Globally Evolutionarily Stable Portfolio Rules. *Journal of Economic Theory*, 140(1):197 – 228, 2008.

Eugene F. Fama. The Behavior of Stock-Market Prices. *Journal of Business*, 38(1):34–105, January 1965.

Eugene F. Fama and Kenneth R. French. Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay? *Journal of Financial Economics*, 60(1):3–43, April 2001.

Milton Friedman. *Essays in Positive Economics.* Chicago: University of Chicago Press, 1953.

Gustavo Grullon, Roni Michaely, and Bhaskaran Swaminathan. Are Dividend Changes a Sign of Firm Maturity? *Journal of Business*, 75 (3):387–424, July 2002.

Bronwyn H. Hall. The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector. *Journal of Industrial Economics*, 35(4):583–606, June 1987.

Richard D. F. Harris and Elias Tzavalis. Testing for Unit Roots in Dynamic Panels in the Presence of a Deterministic Trend: Re-Examining the Unit Root Hypothesis for Real Stock Prices and Dividends. *Econometric Reviews*, 23(2):149–166, January 2004.

Thorsten Hens and Klaus Reiner Schenk-Hoppé. Survival of the Fittest on Wall Street. Working Paper 151, NCCR Finrisk, February 2004.

Thorsten Hens and Klaus Reiner Schenk-Hoppé. Evolutionary Stability of Portfolio Rules in Incomplete Markets. *Journal of Mathematical Economics*, 41(1-2):43–66, February 2005.

Thorsten Hens, Klaus Reiner Schenk-Hoppé, and Martin Stalder. An Application of Evolutionary Finance to Firms Listed in the Swiss Market Index. *Swiss Journal of Economics and Statistics*, 138(4):465–487, 2002.

William J. Hurley and Lewis D. Johnson. A Realistic Dividend Valuation Model. *Financial Analysts Journal*, 50(3):50–54, July/August 1994.

Roger G. Ibbotson and Jeffrey F. Jaffe. "Hot Issue" Markets. *Journal of Finance*, 30(4):1027–1042, September 1975.

John L. Kelly. A New Interpretation of Information Rate. *Bell System Technical Journal*, 35:917–926, July 1956.

Allan W. Kleidon. Variance Bounds Tests and Stock Price Valuation Models. *Journal of Political Economy*, 94(5):953–1001, October 1986.

J. Bradford De Long, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. Positive Feedback Investment Strategies and Destabilizing Rational Speculation. *Journal of Finance*, 45(2):374–97, June 1990.

Dennis C. Mueller. A Life Cycle Theory of the Firm. *Journal of Industrial Economics*, 20(3):199–219, July 1972.

Lubos Pástor and Pietro Veronesi. Rational IPO Waves. *Journal of Finance*, 60(4):1713–1757, August 2005.

Paul A. Samuelson. Evolution and Market Behavior. *Journal of Banking & Finance*, 3(4):305–307, December 1979.

Alvaro Sandroni. Do Markets Favor Agents Able to Make Accurate Predicitions? *Econometrica*, 68(6):1303–1342, November 2000.

Robert J. Shiller. The Use of Volatility Measures in Assessing Market Efficiency. *Journal of Finance*, 36(2):291–304, May 1981.

Andreas Tupak. *Essays in Numerical Evolutionary Finance*. PhD thesis, University of Zurich, 2009.

Maria Vassalou and Yuhang Xing. Default Risk in Equity Returns. *Journal of Finance*, 59(2):831–868, April 2004.

Appendix A

Curriculum Vitae

Name	Urs Schweri
Date of Birth	18. February 1977
Habitat	Siglistorf AG
Email	urs.schweri@uschweri.ch

Education	
2004–2010	University of Zurich, Swiss Finance Institute PhD Program in Finance
1998–2004	University of Zurich, Major in Economics (lic. oec. publ.), *Main emphasis in econometrics*
1994–1998	Kantonsschule at Baden, Maturity Type C, *Main emphasis on science*
1984–1994	Primary School, Sekundarschule and Bezirksschule at Spreitenbach

Professional Experience	
2009–	Zurich Financial Services, Investment Management, Market Strategy and Macroeconomics *Part-time*
2005–	University of Zurich, Chair of Prof. Hens (Financial Economics) *Assistant*
2002–2005	University of Zurich, Chair of Prof. Garbers (Econometrics) *Assistant*
1999–2002	UBS Switzerland, Controlling & Finance, *Internship and later part-time 10 %*
1998	UBS Switzerland, IT, *Internship*